

Detecting and Handling Reflection Symmetries in Mixed-Integer (Nonlinear) Programming

Christopher Hojny¹

¹*Eindhoven University of Technology, Combinatorial Optimization Group, PO Box 513, 5600 MB Eindhoven, The Netherlands, email c.hojny@tue.nl*

May 14, 2024

Abstract

Symmetries in mixed-integer (nonlinear) programs (MINLP), if not handled appropriately, are known to negatively impact the performance of (spatial) branch-and-bound algorithms. Usually one thus tries to remove symmetries from the problem formulation or is relying on a solver that automatically detects and handles symmetries. While modelers of a problem can handle various kinds of symmetries, automatic symmetry detection and handling is mostly restricted to permutation symmetries. This article therefore develops techniques such that also black-box solvers can automatically detect and handle a broader class of symmetries.

Inspired from geometric packing problems such as the kissing number problem, we focus on reflection symmetries of MINLPs. We develop a generic and easily applicable framework that allows to automatically detect reflection symmetries for MINLPs. To handle this broader class of symmetries, we discuss generalizations of state-of-the-art methods for permutation symmetries, and develop dedicated symmetry handling methods for special reflection symmetry groups. Our symmetry detection framework has been implemented in the open-source solver SCIP and we provide a comprehensive discussion of the implementation. The article concludes with a detailed numerical evaluation of our symmetry handling methods when solving MINLPs.

Keywords mixed-integer nonlinear programming • permutation symmetry • reflection symmetry • automatic symmetry detection • automatic symmetry handling

1 Introduction

We consider mixed-integer nonlinear programs (MINLP) of the form

$$\begin{aligned} \min c^\top x \\ g_k(x) \leq 0, \quad k \in \{1, \dots, m\}, \\ \ell_i \leq x_i \leq u_i, \quad i \in \{1, \dots, n\}, \\ x_i \in \mathbb{Z}, \quad i \in \mathcal{I}, \end{aligned} \tag{MINLP}$$

where m and n are positive integers, $c \in \mathbb{R}^n$, $\mathcal{I} \subseteq \{1, \dots, n\}$, and we have, for every $k \in \{1, \dots, m\}$, that $g_k: \mathbb{R}^n \rightarrow \mathbb{R}$ as well as, for each $i \in \{1, \dots, n\}$, that $\ell_i \in \mathbb{R} \cup \{-\infty\}$ and $u_i \in \mathbb{R} \cup \{+\infty\}$. Note that a linear objective $c^\top x$ is without loss of generality as a nonlinear objective $f: \mathbb{R}^n \rightarrow \mathbb{R}$ can be replaced by a single auxiliary variable α and an additional constraint $f(x) - \alpha \leq 0$.

A major technique for solving MINLPs is spatial branch-and-bound [4]. In a nutshell, the idea is to define a relaxation of (MINLP), which is easier to solve than the original problem. This relaxation is then iteratively split into smaller subproblems until an optimal solution of (MINLP) is found or the problem is proven to be infeasible. But if (MINLP) admits symmetries (which will be defined properly below), also the list of subproblems will arguably contain symmetric problems. Branch-and-bound can thus be accelerated by removing symmetric copies from the list. This observation has led to various symmetry handling techniques [5, 9, 13, 14, 16, 19, 21, 24, 25, 28–34, 38,

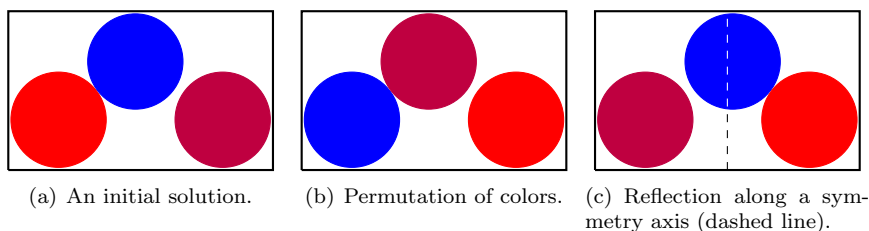


Figure 1: Illustration of permutation and reflection symmetries. The different disks are indicated by colors.

39, 45]. Most of these approaches consider only permutation symmetries, which roughly speaking reorder entries of a solution vector and seem to be most relevant for linear problems. In many MINLPs, however, richer classes of symmetries can arise.

This article’s aim is to extend the literature by tools for reflection symmetries. We believe that this is an important class of symmetries arising in many MINLPs, as illustrated in Example 1.1 below. Such symmetries have, to the best of our knowledge, mainly been handled by reformulating models for specific applications. Our goal is to devise algorithms such that a black-box solver can automatically detect and handle reflection symmetries. We particularly aim for methods that handle more symmetries than model reformulations.

Example 1.1. Let D be a positive integer and let be $B = [-\frac{W}{2}, \frac{W}{2}] \times [-\frac{H}{2}, \frac{H}{2}]$ be a rectangular box of width $W \geq 0$ and height $H \geq 0$. Consider the problem of finding the largest value r such that D non-overlapping disks of radius r can be packed in B . This problem can be modeled as an MINLP, cf. [26, 47]:

$$\begin{aligned}
 & \max r \\
 & (x_i - x_j)^2 + (y_i - y_j)^2 \geq 4r^2, & 1 \leq i < j \leq D, \\
 & -\frac{W}{2} + r \leq x_i \leq \frac{W}{2} - r, & i \in \{1, \dots, D\}, \\
 & -\frac{H}{2} + r \leq y_i \leq \frac{H}{2} - r, & i \in \{1, \dots, D\},
 \end{aligned}$$

where (x_i, y_i) models the center point of disk $i \in \{1, \dots, D\}$.

This problem admits two kinds of symmetries, see Figure 1 for an illustration. First, all disks are equivalent, i.e., for any permutation π of $\{1, \dots, D\}$, replacing (x_i, y_i) by $(x_{\pi(i)}, y_{\pi(i)})$ in a solution of the MINLP leads to an equivalent solution. Second, every solution can be reflected along the reflection symmetry axes of the box B . In formulae, this means one can replace (x_i, y_i) by $(-x_i, y_i)$ for all $i \in \{1, \dots, D\}$ (and similarly for horizontal reflections) to obtain an equivalent solution.

The first class of symmetries corresponds to permutation symmetries, for which detection mechanisms are described in [30, 44]. Except for linear problems [9], however, to the best of our knowledge no general mechanism for detecting reflection symmetries in MINLP has been described. As a consequence, also the literature on methods for handling reflection symmetries in general purpose MINLP solvers is limited. The goals of this article therefore are to

- (G1) provide a framework for detecting both permutation and reflection symmetries in MINLP;
- (G2) develop an easily extendable open-source tool for detecting permutation and reflection symmetries;
- (G3) derive symmetry handling methods for reflection symmetries;
- (G4) evaluate how frequently reflection symmetries arise in MINLP and by how much our methods accelerate the solving process.

The main focus will be on achieving Goal (G2). In contrast to existing symmetry detection frameworks, (G2) requires a flexible mechanism to encode symmetries of an MINLP that can easily incorporate custom constraints whose logic is not known to an MINLP solver. We therefore need to develop an abstract symmetry detection framework for achieving both (G1) and (G2).

To achieve our goals, we proceed as follows. A detailed description of the problem and underlying mathematical concepts is provided in Section 2. To achieve (G1), we provide a detailed description of existing symmetry detection schemes in Section 3. A common abstraction of these schemes is provided in Section 4, and Sections 4.1 and 4.2 describe two instantiations of our abstract framework to detect reflection symmetries in MINLP. Turning to Goal (G2), Section 5 describes an implementation of our abstract framework in the open-source MINLP solver SCIP [10]. Our implementation is included in SCIP since version 9.0, and thus, publicly available. Section 6 provides an overview of existing symmetry handling methods. We also describe how these methods can be adapted for reflection symmetries, cf. (G3). We conclude the article in Section 7 with a detailed analysis of numerical experiments to evaluate the impact of handling reflection symmetries in MINLP, cf. (G4). For a literature review of symmetry detection and handling, we refer the reader to Section 3 and 6, respectively.

2 Problem Statement and Basic Definitions

The arguably most general symmetry of (MINLP) is a map $\sigma: \mathbb{R}^n \rightarrow \mathbb{R}^n$ such that, for all $x \in \mathbb{R}^n$, one has (S1) $\sigma(x)$ is feasible for (MINLP) if and only if x is feasible, and (S2) $c^\top \sigma(x) = c^\top x$. Towards the detection of symmetries, however, this definition might be too general as it is unclear how a general detection mechanism could look like. One therefore usually restricts to subclasses of symmetries, the most popular class being permutation symmetries.

Let n be a positive integer and define $[n] := \{1, \dots, n\}$. Any bijective map $\pi: [n] \rightarrow [n]$ is called a *permutation* of $[n]$. The set of all permutations of $[n]$ is denoted \mathcal{S}_n and forms a group w.r.t. composition, the *symmetric group*. A permutation $\pi \in \mathcal{S}_n$ naturally acts on \mathbb{R}^n as $\pi(x) = (x_{\pi^{-1}(1)}, \dots, x_{\pi^{-1}(n)})$, i.e., π permutes the coordinates of a vector $x \in \mathbb{R}^n$. Note that the inverses are necessary to ensure that this indeed defines a group action on \mathbb{R}^n . We say that $\pi \in \mathcal{S}_n$ is a *permutation symmetry* of (MINLP) if it satisfies (S1) and (S2). The group consisting of all permutation symmetries of (MINLP) is called its *permutation symmetry group*.

Despite being the most popular class of symmetries in the mixed-integer (nonlinear) programming literature, computing all permutation symmetries is NP-hard already for binary linear programs [35]. For general MINLP, it is even undecidable if π defines a permutation symmetry, cf. [30]. For integer programs, the reason is that symmetries are defined based on the feasible region of (MINLP), which is not known explicitly, and for MINLP the difficulty arises that determining feasibility of a set of linear equations is undecidable [49]. A possible remedy is thus to restrict to permutation symmetries that keep a particular formulation of (MINLP) invariant. We discuss this approach in more detail in Section 3 and turn to the definition of reflection symmetries next.

The basis for our definition of reflection symmetries are signed permutations. We call a map $\gamma: \{\pm 1, \dots, \pm n\} \rightarrow \{\pm 1, \dots, \pm n\}$ *signed permutation* if it is bijective and satisfies $\gamma(-i) = -\gamma(i)$ for every $i \in \{\pm 1, \dots, \pm n\}$. Signed permutations form a group under composition, the *signed symmetric group*, denoted \mathcal{S}_n^\pm . Analogously to permutations, we define a group action on \mathbb{R}^n via

$$\gamma(x) = (\text{sgn}(\gamma^{-1}(1))x_{|\gamma^{-1}(1)|}, \dots, \text{sgn}(\gamma^{-1}(n))x_{|\gamma^{-1}(n)|}),$$

where $\text{sgn}: \mathbb{R} \rightarrow \{0, \pm 1\}$ denotes the sign operator. That is, due to taking the absolute values in the indices, γ reorders the entries of x like a permutation, but can also change the sign of some entries. A signed permutation thus reorders the entries of a vector and reflects it along some of the standard hyperplanes with normal vectors being the standard unit vectors.

Example 2.1 (Example 1.1 continued). Consider the situation of Example 1.1 with three disks as illustrated in Figure 1. Collect all variables in a common vector $z = (x_1, x_2, x_3, y_1, y_2, y_3, r)$, where (x_1, y_1) , (x_2, y_2) , and (x_3, y_3) corresponds to the red, blue, and purple disk, respectively. The permutation symmetry of Figure 1(b) corresponds to $\pi = (1, 3, 2)(4, 6, 5)$, where the cycle

notation denotes that z_1 is mapped to z_3 , z_3 is mapped to z_2 , and z_2 is mapped to z_1 etc. Indeed, $\pi(z) = (z_{\pi^{-1}(1)}, \dots, z_{\pi^{-1}(7)})$ is given by

$$\pi(z) = (z_3, z_2, z_1, z_6, z_5, z_4, z_7) = (x_3, x_2, x_1, y_3, y_2, y_1, r),$$

which corresponds to the exchange of colored disks. The reflection symmetry of Figure 1(c) corresponds to the signed permutation $\gamma = (1, -1)(2, -2)(3, -3)$, where the cycle notation now also takes signs into account. Then, symmetric solution $\gamma(z) = (-x_1, -x_2, -x_3, y_1, y_2, y_3, r)$ indeed reflects all x -coordinates.

While signed permutations capture reflections along standard hyperplanes, they do not allow for reflections along general affine hyperplanes with standard normal vectors. For example, if the box in Example 1.1 is not centered at the origin, but is given by $[0, W] \times [0, H]$, no signed permutation can express the reflection $(x, y) \mapsto (W - x, y)$ along $x = \frac{W}{2}$. To also capture such symmetries, we introduce another action of $\gamma \in \mathcal{S}_n^\pm$ on \mathbb{R}^n that is adapted to an MINLP.

We still propose to encode our more general reflection symmetries via signed permutations. But instead of necessarily reflecting a variable x_i at the origin, we reflect a variable at the center of its domain $\ell_i \leq x_i \leq u_i$ as given by the bounds of (MINLP). Note, however, that the center is not well-defined in case $[\ell_i, u_i]$ defines a half-open interval. We therefore introduce the set

$$\mathcal{C} := \{i \in [n] : |\ell_i| = u_i = \infty \text{ or both } \ell_i \text{ and } u_i \text{ are finite}\}$$

containing the indices of variables whose domain has a well-defined center. For $i \in \mathcal{C}$, the *center* of variable x_i is $\xi_i = \frac{\ell_i + u_i}{2}$, where we define $-\infty + \infty = 0$. For $i \notin \mathcal{C}$, we define $\xi_i = 0$. With this notion, the reflection of variable x_i , $i \in [n]$, is given by $x_i \mapsto 2\xi_i - x_i$. If $i \in \mathcal{C}$, variable x_i is reflected at its domain center; in particular, a variable can be reflection-symmetric to itself. For $i \in [n] \setminus \mathcal{C}$, a variable is reflected at the origin.

Definition 2.2. Consider (MINLP) with lower and upper bounds ℓ and u , respectively, used to derive \mathcal{C} and the corresponding center points. For $\gamma \in \mathcal{S}_n^\pm$, the *reflection* $\rho: \mathbb{R}^n \rightarrow \mathbb{R}^n$, for $i \in [n]$, is

$$\rho(x; \gamma)_i = \xi_i + \text{sgn}(\gamma^{-1}(i)) \cdot (x_{|\gamma^{-1}(i)|} - \xi_{|\gamma^{-1}(i)|}).$$

That is, if $\text{sgn}(\gamma^{-1}(i)) < 0$, then the pre-image of $\rho(x; \gamma)_i$ is reflected according to the map $x_{|\gamma^{-1}(i)|} \mapsto 2\xi_{|\gamma^{-1}(i)|} - x_{|\gamma^{-1}(i)|}$ and afterwards centered around ξ_i via the translation $\xi_i - \xi_{|\gamma^{-1}(i)|}$. Otherwise, if $\text{sgn}(\gamma^{-1}(i)) > 0$, $\rho(x; \gamma)_i$ arises by an ordinary permutation and applying the same translation as before.

Definition 2.3. A signed permutation $\gamma \in \mathcal{S}_n^\pm$ defines a *reflection symmetry* of (MINLP) if $\rho(\cdot; \gamma)$ satisfies (S1) and (S2).

Our aim is to achieve Goals (G1)–(G4) for reflection symmetries of (MINLP).

We close this section by showing that reflections define an action on \mathbb{R}^n , i.e., for $\gamma_1, \gamma_2 \in \mathcal{S}_n^\pm$, their composition $\gamma_2 \circ \gamma_1$ satisfies $\rho(x; \gamma_2 \circ \gamma_1) = \rho(\rho(x; \gamma_1); \gamma_2)$, and, if id denotes the identity, then $\rho(x; \text{id}) = x$. Thus, our definition of reflections is independent from whether one first composes signed permutations and applies a single reflection or whether one applies a series of reflections for the corresponding signed permutations.

Lemma 2.4. Let $\Gamma \leq \mathcal{S}_n^\pm$ and let ξ_i , $i \in [n]$, be the reflection center for the i -th coordinate. Then, $\rho(\cdot; \gamma)$, $\gamma \in \Gamma$, defines a group action on \mathbb{R}^n .

Proof. Let $x \in \mathbb{R}^n$. Then, $\rho(x; \text{id}) = x$ follows immediately. It thus remains to show for $\gamma_1, \gamma_2 \in \mathcal{S}_n^\pm$ that $\rho(x; \gamma_2 \circ \gamma_1) = \rho(\rho(x; \gamma_1); \gamma_2)$. To this end, observe that, for any $i \in [n]$, we have

$$|\gamma_1^{-1}(|\gamma_2^{-1}(i)|)| = |(\gamma_1^{-1} \circ \gamma_2^{-1})(i)| = |(\gamma_2 \circ \gamma_1)^{-1}(i)| \quad (1)$$

as signed permutations γ satisfy $\gamma(-j) = -\gamma(j)$ for all $j \in [n]$. The same relation also implies

$$\text{sgn}(\gamma_2^{-1}(i)) \cdot \text{sgn}(\gamma_1^{-1}(|\gamma_2^{-1}(i)|)) = \text{sgn}(\gamma_1^{-1} \circ \gamma_2^{-1}(i)) = \text{sgn}((\gamma_2 \circ \gamma_1)^{-1}(i)) \quad (2)$$

since $y = \text{sgn}(y) \cdot |y|$ for every $y \in \mathbb{R}$.

Let $\gamma = \gamma_2 \circ \gamma_1$. Then,

$$\begin{aligned} \rho(x; \gamma_1)_{|\gamma_2^{-1}(i)|} &= \xi_{|\gamma_2^{-1}(i)|} + \operatorname{sgn}(\gamma_1^{-1}(|\gamma_2^{-1}(i)|)) \cdot (x_{|\gamma_1^{-1}(|\gamma_2^{-1}(i)|)} - \xi_{|\gamma_1^{-1}(|\gamma_2^{-1}(i)|)}) \\ &\stackrel{(1)}{=} \xi_{|\gamma_2^{-1}(i)|} + \operatorname{sgn}(\gamma_1^{-1}(|\gamma_2^{-1}(i)|)) \cdot (x_{|\gamma^{-1}(i)|} - \xi_{|\gamma^{-1}(i)|}), \end{aligned}$$

and thus,

$$\begin{aligned} \rho(\rho(x; \gamma_1); \gamma_2)_i &= \xi_i + \operatorname{sgn}(\gamma_2^{-1}(i)) \cdot (\rho(x; \gamma_1)_{|\gamma_2^{-1}(i)|} - \xi_{|\gamma_2^{-1}(i)|}) \\ &= \xi_i + \operatorname{sgn}(\gamma_2^{-1}(i)) \cdot (\xi_{|\gamma_2^{-1}(i)|} - \xi_{|\gamma_2^{-1}(i)|}) \\ &\quad + \operatorname{sgn}(\gamma_2^{-1}(i)) \cdot \operatorname{sgn}(\gamma_1^{-1}(|\gamma_2^{-1}(i)|)) \cdot (x_{|\gamma^{-1}(i)|} - \xi_{|\gamma^{-1}(i)|}) \\ &\stackrel{(2)}{=} \xi_i + \operatorname{sgn}(\gamma^{-1}(i)) \cdot (x_{|\gamma^{-1}(i)|} - \xi_{|\gamma^{-1}(i)|}) \\ &= \rho(x; \gamma). \end{aligned}$$

Thus, $\rho(\cdot; \gamma)$ defines a group action. \square

3 Existing Symmetry Detection Schemes

In preparation for our symmetry detection framework, this section reviews four schemes discussed in the literature. These schemes have been developed to detect permutation symmetries of mixed-integer linear programs (MILP), signed permutation symmetries of MILPs, reflection symmetries in satisfiability problems (SAT), and permutation symmetries of MINLPs, respectively.

Permutation Symmetries of MILP. Since already deciding whether \mathcal{S}_n is the permutation symmetry group of an MILP is NP-complete [35], one usually only considers symmetries that keep the formulation of an MILP invariant. Given an MILP

$$\min\{c^\top x : Ax \leq b, x_i \in \mathbb{Z} \text{ for all } i \in \mathcal{I}, x \in \mathbb{R}^n\},$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $\mathcal{I} \subseteq [n]$, a *formulation symmetry* is a permutation $\pi \in \mathcal{S}_n$ for which there exists $\pi' \in \mathcal{S}_m$ such that

- $\pi(c) = c$ and $\pi'(b) = b$;
- for all $(i, j) \in [m] \times [n]$, one has $A_{\pi'(i), \pi(j)} = A_{i, j}$;
- π keeps \mathcal{I} invariant.

For detecting formulation symmetries, [44] observes that the formulation symmetry group of an MILP is isomorphic to the color-preserving automorphism group of the following graph. This graph contains a node for every variable as well as each constraint from $Ax \leq b$. Every variable node receives a color, which is determined based on the variable's type (objective coefficient, an (non-) integrality); every constraint node gets a color based on the constraint's right-hand side. Moreover, there is an edge between the node of variable x_j and the node of the i -th constraint if and only if $A_{i, j} \neq 0$. The edge is colored based on the coefficient $A_{i, j}$. An illustration of the symmetry detection graph for the MILP

$$\begin{aligned} \min \quad & x_1 - x_2 + 2x_3 + 2x_4 \\ & x_3 + x_4 \leq 1 \\ & -x_1 + x_2 + 3x_3 \leq 4 \\ & -x_1 + x_2 + 3x_4 \leq 4 \end{aligned} \tag{3}$$

is given in Figure 2(a).

Automorphisms of such symmetry detection graphs can be found by graph automorphism tools such as BLISS [22], DEJAVU [1], or NAUTY [36]. We remark that these tools can only handle node colors, but not edge colors. The graphs thus need to be manipulated by replacing an edge $\{u, v\}$ by an auxiliary node w , which receives the color of $\{u, v\}$, and the edges $\{u, w\}$ and $\{v, w\}$. To reduce the number of auxiliary nodes, grouping techniques can identify auxiliary nodes of several edges with each other, see [40] and Section 5.3.

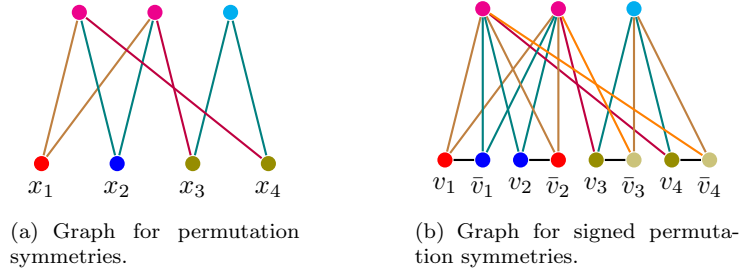


Figure 2: Illustration of symmetry detection graphs for Problem (3).

Signed Permutation Symmetries of MILP. For detecting formulation symmetries of an MILP corresponding to signed permutations, [9] also suggests to find automorphisms of a colored graph. For every variable x_j , their graph contains two nodes v_j and \bar{v}_j , where v_j will represent the original variable x_j and \bar{v}_j the negation $-x_j$. Moreover, the i -th constraint is represented by a node w_i , and every entry $A_{i,j}$ of the constraint A gets two nodes $a_{i,j}$ and $\bar{a}_{i,j}$. Finally, there are three groups of nodes representing the numerical coefficients of the MILP: one group consisting of the right-hand side values $\{b_1, \dots, b_m\}$, another group consisting of the original and negated objective coefficients $\{\pm c_1, \dots, \pm c_n\}$, and the last group consisting of the original and negated matrix entries $\{\pm A_{i,j} : (i, j) \in [m] \times [n]\}$. The edge set is given by

$$\begin{aligned}
& \{\{v_j, c_j\}, \{\bar{v}_j, -c_j\} : j \in [n]\} \\
& \cup \{\{w_i, b_i\} : i \in [m]\} \\
& \cup \{\{a_{i,j}, A_{i,j}\}, \{v_j, a_{i,j}\}, \{w_i, a_{i,j}\} : (i, j) \in [m] \times [n]\} \\
& \cup \{\{\bar{a}_{i,j}, -A_{i,j}\}, \{\bar{v}_j, \bar{a}_{i,j}\}, \{w_i, \bar{a}_{i,j}\} : (i, j) \in [m] \times [n]\} \\
& \cup \{\{v_j, \bar{v}_j\} : j \in [n]\}
\end{aligned}$$

In contrast to [44], edges remain uncolored. Instead, there is a unique color for each of the variable, constraint, and coefficients groups $\{v_j, \bar{v}_j : j \in [n]\}$, $\{w_i : i \in [n]\}$, and $\{a_{i,j}, \bar{a}_{i,j} : (i, j) \in [m] \times [n]\}$, respectively; nodes representing a numerical value, receive a color corresponding to their numerical value.

The idea of this graph is to not only encode the original constraint system as in [44], but also the negated constraint system. The last class of edges $\{v_j, \bar{v}_j\}$, $j \in [n]$, ensures the property that $\gamma(-j) = -\gamma(j)$ for a signed permutation γ .

Note that the graph can be represented more compactly: On the one hand, numerical value nodes can be removed and instead nodes adjacent with them can receive their color. On the other hand, nodes $a_{i,j}$ and $\bar{a}_{i,j}$ can be removed, and instead one connects v_j and w_i by an edge colored according to $A_{i,j}$ (analogously for negated variable nodes \bar{v}_j). One can thus interpret nodes $a_{i,j}$ and $\bar{a}_{i,j}$ as the auxiliary nodes that arise when substituting colored edges in the graph of [44]. As such, the grouping techniques described in [40] can also be used to reduce the number of needed nodes $a_{i,j}$ and $\bar{a}_{i,j}$. Figure 2(b) illustrates the more compact graph, better revealing relations to the graph by [44].

Reflection Symmetries for SAT Consider variables $x_1, \dots, x_n \in \{0, 1\}$. The reflection of x_j is given by $\bar{x}_j = 1 - x_j$, cf. Definition 2.2. Using MINLP notation, a SAT formula can be represented by a set of m linear constraints

$$\sum_{j \in J_i^+} x_j + \sum_{j \in J_i^-} \bar{x}_j \geq 1,$$

where $J_i^+, J_i^- \subseteq [n]$ are disjoint for every $i \in [m]$, and the objective being constant 0. That is, SAT problems are pure feasibility problems.

A common way to detect reflection symmetries for SAT is described in [43]. As in [9], there are nodes v_j and \bar{v}_j to represent variable x_j and its reflection \bar{x}_j for $j \in [n]$, and there is a node w_i

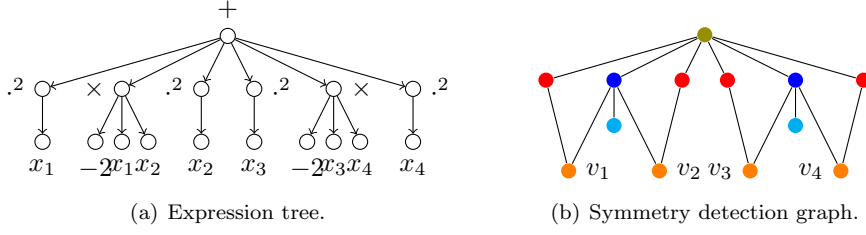


Figure 3: Expression tree and symmetry detection graph for $x_1^2 - 2x_1x_2 + x_2^2 + x_3^2 - 2x_3x_4 + x_4^2$.

for every $i \in [m]$; the edge set is

$$\{\{w_i, v_j\} : i \in [m], j \in J_i^+\} \cup \{\{w_i, \bar{v}_j\} : i \in [m], j \in J_i^-\} \cup \{\{v_j, \bar{v}_j\} : j \in [n]\}.$$

The role of the last group of edges is again to make sure that if a symmetry maps v_j onto $v_{j'}$ or $\bar{v}_{j'}$, then $\bar{v}_{j'}$ is mapped onto \bar{v}_j or v_j , respectively.

In contrast to the previous graphs, no colors are needed since there is only one type of coefficients. Furthermore, this graph allows to detect reflection symmetries, because all variables have the same domain. Despite the similarities to the previous graphs for MILP, the latter cannot immediately be used for detecting reflection symmetries in MILP as variables might have different domains. We resolve this issue in our symmetry detection scheme in Section 4.

Permutation Symmetries in MINLP As discussed above, one of the difficulties for detecting symmetries of (MINLP) is that checking whether two nonlinear equations are equivalent is undecidable. The symmetry detection scheme of [30] resolves this issue by restricting to nonlinear constraints $g_k(x)$, $k \in [m]$, that admit a representation via expression trees [11]. An *expression tree* for g_k is an arborescence T_k in which all arcs point away from the root node. Each leaf node belongs to exactly one of two classes, it is either a variable or value node, and the non-leaf nodes are operator nodes. Each variable node corresponds to a variable present in g_k and each value node holds a numerical value. Every operator node v corresponds to an d -ary mathematical operator, where d is the number of children of v . The nonlinear function g_k can then be recovered from T_k by iteratively evaluating the nodes of T_k , where evaluating a node means to assign it a mathematical function. Leaf nodes evaluate to the corresponding variables or numerical values. Operator nodes v are evaluated by applying the operator of v to the functions assigned to the children of v . By evaluating the nodes in a bottom-up fashion, the last processed node is the root node, whose assigned function is g_k , see Figure 3(a) for an illustration.

Expression trees can also make use of non-commutative operators, e.g., the minus-operator. In this case, arcs receive labels to indicate the order of the input. For the ease of exposition, we assume all operators to be commutative, but all results can be easily refined for non-commutative operators.

To detect permutation symmetries for MINLPs whose constraints are represented via expression trees, [30] suggests to derive a symmetry detection graph from the expression trees, see Figure 3(b) for an illustration. For each variable x_j , $j \in [n]$, construct a node v_j , and, for each constraint g_k , $k \in [m]$, of (MINLP), construct an undirected version of T_k . Remove the leaves corresponding to variables x_j and connect their parents to v_j instead. That is, the entire graph contains only a single node corresponding to x_j . Then, the operators and numerical values are interpreted as distinct colors. Finally, color the nodes v_j , $j \in [n]$, according to their type (same objective coefficient, and lower and upper bounds).¹ One can show that every color preserving automorphism of this graph corresponds to a symmetry of (MINLP).

¹Formally, [30] introduces a separate expression tree for the objective, but due to our assumption of a linear objective, the graph we describe is a bit simpler.

4 A Framework for Symmetry Detection in MINLP

The core of our symmetry detection framework is an abstract notion of a symmetry detection graph, generalizing the existing frameworks. This will be useful, on the one hand, since different classes of constraints (e.g., linear, nonlinear) might allow for more compact representations to detect symmetries. On the other hand, this notion is flexible towards achieving Goal (G2), the development of an easily extendable open-source tool. After defining symmetry detection graphs, Section 4.1 provides a basic instantiation of the abstract framework to detect reflection symmetries in MINLPs. Section 4.2 then refines the previous concepts to detect more symmetries.

Before we define symmetry detection graphs, we introduce some notation. Consider a list of variables x_i , $i \in [n]$, with lower bounds $\ell_i \in \mathbb{R} \cup \{-\infty\}$ and upper bounds $u_i \in \mathbb{R} \cup \{\infty\}$, as well as objective coefficient $c_i \in \mathbb{R}$. For $i \in [n]$, let ξ_i be the reflection center of variable x_i as defined in Section 2, and denote by x_{-i} the reflection of x_i at ξ_i . That is, variable x_{-i} has lower bound $\ell_{-i} = 2\xi_i - u_i$, upper bound $u_{-i} = 2\xi_i - \ell_i$, and objective coefficient $c_{-i} = -c_i$. Moreover, for a positive integer n , let $[\pm n] := \{\pm 1, \dots, \pm n\}$.

Definition 4.1. Let P be an instance of (MINLP). Let G be a node and edge colored connected graph that contains pairwise distinct distinguished nodes v_i , $i \in [\pm n]$. We call G a *symmetry detection graph (SDG)* for P if each color-preserving automorphism π of G satisfies:

1. π keeps $\{v_i : i \in [\pm n]\}$ invariant;
2. if $\pi(v_i) = v_j$ for some $i, j \in [\pm n]$, then $\pi(v_{-i}) = v_{-j}$;
3. the signed permutation $\gamma \in \mathcal{S}_n^\pm$ defines a reflection symmetry of P , where, for each $i \in [\pm n]$, $\gamma(i) = j$ for the unique j with $\pi(v_i) = v_j$.

Note that SDGs are well-defined: Due to the first property, the second property is well-defined, and γ in the third property is indeed a signed permutation due to the first and second property.

Definition 4.1 seems to be the right notion for detecting reflection symmetries as it provides an abstract generalization of the symmetry detection frameworks of Section 3. But to use it, one needs a concrete mechanism to build an SDG. While black-box solvers can use fixed sets of rules for building SDGs, an easily extendable tool needs to be flexible to also support unknown custom constraints, e.g., nonlinear constraints not admitting a (simple) representation via expression trees. We therefore suggest to use an abstract mechanism to build an SDG from SDGs of single constraints as formalized next. This mechanism is the core of our implementation, which we discuss in Section 5 in detail.

Let P be an instance of (MINLP) with m constraints $g_1, \dots, g_m: \mathbb{R}^n \rightarrow \mathbb{R}$. For $k \in [m]$, denote by P_k the MINLP arising from P by removing all constraints except for g_k . To build an SDG for P from the SDGs G_1, \dots, G_m of P_1, \dots, P_m , our idea is to take the disjoint union of G_1, \dots, G_m , and to identify the distinguished nodes for variable x_i , $i \in [\pm n]$, with each other. To apply this idea, we need to make sure that the colors of the different SDGs match. That is, equivalent objects can only be mapped to equivalent objects.

To model equivalence of variables, we introduce variable types. For a variable x_i , $i \in [n]$, of (MINLP) with lower bound ℓ_i , upper bound u_i , objective coefficient c_i , and a Boolean encoding if $i \in \mathcal{I}$, the *type* of x_i is

$$t(x_i) := (\ell_i - \xi_i, u_i - \xi_i, c_i, i \in \mathcal{I}).$$

That is, the type is defined by the lower and upper bounds relative to the reflection center, the objective coefficient, and integrality status. Using relative lower and upper bounds instead of absolute bounds will allow us to detect symmetries also of variables with different domain centers. The type of x_{-i} is defined based on the reflected variable, i.e., $t(x_{-i}) = (\xi_i - u_i, \xi_i - \ell_i, -c_i, i \in \mathcal{I})$. We then define the set of variable colors as $\mathfrak{V} = \{t(x_i) : i \in [\pm n]\}$, i.e., each type is associated with a unique color. An SDG is called *variable color compatible* if node v_i associated with variable x_i , $i \in [\pm n]$, is colored by color $t(x_i)$ and none of the non-distinguished nodes receives a color from \mathfrak{V} .

Besides mapping variables of the same type to each other, we need to ensure that the set of constraints remains invariant after applying a reflection symmetry. This is achieved via the concepts of anchors and constraint compatibility. An SDG $G = (V, E)$ is called *anchored* if there is a non-distinguished $a \in V$ such that, for each $v \in V \setminus \{a\}$, there is a path from a to v in G such that

neither of the interior nodes along the path is a distinguished variable node. We call a an *anchor* of G . For the second concept, let $k, k' \in [m]$. Let $G_k = (V_k, E_k)$ and $G_{k'} = (V_{k'}, E_{k'})$ be SDGs for P_k and $P_{k'}$, respectively. We call G_k and $G_{k'}$ *constraint compatible* if they are non-isomorphic or $g_k \equiv g_{k'}$, where $g_k \equiv g_{k'}$ means that there is a permutation $\pi \in \mathcal{S}_n$ such that $g_k(x) = g_{k'}(\pi(x))$.

Theorem 4.2. *Let P be an MINLP with m constraints. Let $G_k = (V_k, E_k)$, $k \in [m]$, be an anchored SDG for P_k with distinguished nodes v_i^k , $i \in [\pm n]$. Suppose G_1, \dots, G_m are variable color and constraint compatible. Let v_i , $i \in [\pm n]$, be pairwise distinct with color $t(x_i)$ and let*

$$\begin{aligned} V &= \{v_i : i \in [\pm n]\} \cup \bigcup_{k=1}^m V_k \setminus \{v_i^k : i \in [\pm n]\}, \\ E &= \{\{v_i, v_{-i}\} : i \in [n]\} \cup \bigcup_{k=1}^m \{\{v_i, v\} : \{v_i^k, v\} \in E_k \text{ and } v \neq v_{-i}^k\} \\ &\quad \cup \bigcup_{k=1}^m \{e \in E_k : e \text{ contains no distinguished node of } G_k\}. \end{aligned}$$

Then, $G = (V, E)$ is an SDG for P with distinguished nodes v_i , $i \in [\pm n]$.

Proof. Let π be a color-preserving automorphism of G . We need to show that π satisfies the three properties of Definition 4.1. The first property holds since v_i , $i \in [\pm n]$, are the only nodes receiving a variable color from \mathfrak{A} by variable color compatibility of G_1, \dots, G_m . The second property is satisfied, because the only edges between the distinguished nodes v_i , $i \in [\pm n]$, are the edges $\{v_i, v_{-i}\}$. Consequently, if $\pi(v_i) = v_j$ for some $j \in [\pm n]$, also $\pi(v_{-i}) = v_{-j}$ holds.

For verifying the last property, observe that each path in G corresponds to a path in $\pi(G)$. This in particular holds for the paths originating from the anchors of G_1, \dots, G_m to the remaining nodes of the respective graphs. Consequently, if π maps a non-distinguished node of G_k to a non-distinguished node of $G_{k'}$ for $k, k' \in [m]$, it also maps the remaining non-distinguished nodes of G_k to nodes of $G_{k'}$. As a consequence, G_k and $G_{k'}$ are isomorphic, because we can identify v_i , $i \in [\pm n]$, with both v_i^k and $v_i^{k'}$. By constraint compatibility of G_1, \dots, G_m we thus conclude $g_k \equiv g_{k'}$. Property three for G then follows from the same property for G_1, \dots, G_m since we can mutually identify the distinguished nodes of G_1, \dots, G_m with the distinguished nodes of G . \square

Note that all frameworks of Section 3 are (slight variations of) an instantiation of Theorem 4.2. When ignoring colors, the framework of [43] for SAT problems follows Theorem 4.2, and when reflecting variables at the origin, the framework of [9] for signed permutations for MILPs can be derived from Theorem 4.2. If one ignores reflections and has only distinguished nodes for non-reflected variables, also the frameworks for MILP [44] and MINLP [30] follow from Theorem 4.2.

4.1 Basic Framework for Detecting Reflection Symmetries in MINLP

We now turn the focus to the detection of reflection symmetries of an instance P of (MINLP) with m constraints that are represented by expression trees. Due to Theorem 4.2, it is sufficient to find, for each $k \in [m]$, an anchored SDG G_k for P_k such that G_1, \dots, G_m are variable color and constraint compatible. To this end, we combine the ideas of [9] and [30]. We will illustrate the concepts introduced in this section using the example

$$\min\{0 : 4x_1 - 4x_2 + x_3 - x_4 \leq 0, x_1, x_2 \in [-1, 1], x_3 \in [1, 3], x_4 \in [-2, 0]\}. \quad (4)$$

For deriving our SDG, it will be convenient to apply two modifications to an expression tree T . First, we transform (MINLP) such that all reflection centers ξ_i , $i \in [n]$, are at 0. This can be achieved by replacing all variables x_i by $x_i - \xi_i$, as already indicated by the relative bounds of variable type $t(x_i)$. To still represent the original MINLP, we apply two modifications: (i) set the lower and upper bounds of x_i to $\ell_i - \xi_i$ and $u_i - \xi_i$, respectively, and (ii) replace in T every sub expression tree representing an expression $\alpha \cdot x_i$ for some $\alpha \in \mathbb{R}$ by an expression tree for the expression $\alpha \cdot x_i + \alpha \cdot \xi_i$. Second, observe that reflecting the expression $\alpha \cdot x_i$ for some $\alpha \in \mathbb{R}$, results in $\alpha \cdot (2\xi_i - x_i)$. That is, the sign of the coefficient of x_i changes. To easily model this in an SDG,

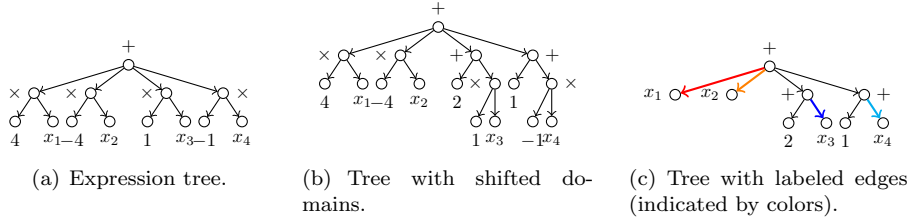


Figure 4: Illustration of the two preprocessing modifications of expression tree for (4).

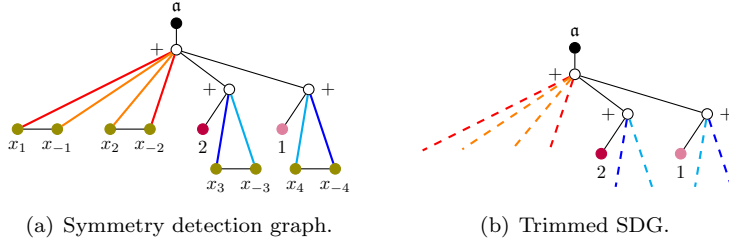


Figure 5: Illustration of an SDG and the corresponding trimmed SDG for (4).

we encode variable coefficients as edge colors. More concretely, suppose T has an operator node v that corresponds to a multiplication operation and that has exactly two children. Whenever one child v_1 is a variable and the other child v_2 is a numerical value, we remove v_2 from T and assign its value to the arc connecting v and v_1 , see Figure 4 for an illustration of both operations.

In the following, we assume that all expression trees incorporate the two previously described modifications. To construct colors for SDGs, let \mathfrak{N} be the set of all *numerical values* used in the expression trees of MINLP P , and define $\mathfrak{N} = \{\pm v : v \in \mathfrak{N}'\}$. Let \mathfrak{D} be the set of all *operators* appearing in the expression trees, and let \mathbf{a} be a symbol representing the anchor of a graph. Moreover, recall that \mathfrak{V} denotes the set of all types of variables in P . Then, $\mathfrak{C} = \mathfrak{N} \cup \mathfrak{D} \cup \mathfrak{V} \cup \{\mathbf{a}\}$ contains all types of nodes needed to build an SDG. We refer to \mathfrak{C} as *colors*, i.e., we associate a color to each node based on its type.

Our construction of an SDG closely follows [30]. We derive an undirected copy of the expression tree of constraint k , denoted $G'_k = (V'_k, E'_k)$. Each node in V'_k receives the color from \mathfrak{C} corresponding to its type. If an arc has received a value due to preprocessing, we color the corresponding edge in E'_k by the corresponding color in \mathfrak{C} . Each uncolored edge incident with a variable node receives the color of the numerical value 1. Finally, we add an anchor node a to G'_k and replace nodes corresponding to variables by a gadget. The anchor node a receives color \mathbf{a} and is connected via an edge with the root node of the expression tree. To represent variable x_i , $i \in [n]$, we introduce two distinguished nodes v_i^k and v_{-i}^k that are colored by $t(x_i)$ and $t(x_{-i})$, respectively. Both nodes are connected by an edge $\{v_i^k, v_{-i}^k\}$. Furthermore, every edge $\{v, v_i\} \in E'_k$ with color $c \in \mathfrak{N}$ that is incident to a node v_i representing variable x_i is replaced by two edges $\{v, v_i^k\}$ and $\{v, v_{-i}^k\}$ with color c and $-c$, respectively. We denote the resulting graph by G_k , see Figure 5(a) for an illustration. Note that all variable nodes are equally colored as all shifted variable domains are identical in our example.

Proposition 4.3. *Let P be an MINLP with m constraints represented by expression trees T_1, \dots, T_m all of whose operators are commutative. For $k \in [m]$, let $G_k = (V_k, E_k)$ be the colored graph derived from T_k as described above. Then, for every $k \in [m]$, G_k is an anchored SDG for P_k . Moreover, G_1, \dots, G_m are variable color and constraint compatible.*

Proof. For every $k \in [m]$, graph G_k is anchored since each variable node of T_k is a leaf and the anchor node of G_k is connected with the root of T_k . Moreover, G_1, \dots, G_m are constraint compatible since they are essentially copies of the corresponding expression trees. The graphs G_1, \dots, G_m are also variable color compatible, because variable nodes are colored according to their type. It thus suffices to show that each graph G_k , $k \in [m]$, is an SDG.

The first two properties of SDGs follow immediately from coloring the distinguished nodes by the corresponding variables' types and the only edges connecting the distinguished nodes being $\{v_i^k, v_{-i}^k\}$, $i \in [n]$. Regarding the third property, observe that expression tree T_k is almost identical to G_k . The main differences are (i) the anchor node in G_k has no counterpart in T_k ; (ii) there are possibly multiple copies of variable nodes in T_k , whereas these nodes are identified in G_k via the distinguished nodes; (iii) G_k also contains nodes for x_{-1}, \dots, x_{-n} . We now argue that these differences ensure that G_k is an SDG. To this end, let \bar{T}_k be the graph arising from T_k by removing all variable nodes, see Figure 5(b). As for expression trees, we can associate with \bar{T}_k an arithmetic expression in which the input of some operators is not fully specified. We refer to such an expression as a *pre-function*.

Let π be an automorphism of G_k and let γ be the signed permutation associated with π as defined in Definition 4.1. Let $\bar{\pi}$ be the restriction of π onto \bar{T}_k . Then, $\bar{\pi}$ is well-defined since every node in \bar{T}_k has a unique counterpart in G_k (there are no variable nodes in \bar{T}_k). We can thus interpret \bar{T}_k as an induced subgraph of G_k . Since the anchor of G_k is the only node colored by \mathbf{a} and the anchor is only connected with T_k 's root, $\bar{\pi}$ is an automorphism of \bar{T}_k that keeps the root invariant. The pre-functions associated with $\pi(\bar{T}_k)$ and \bar{T}_k are hence the same. To conclude the proof, we show that inserting x and $\gamma(x)$ into the pre-functions for \bar{T}_k and $\pi(\bar{T}_k)$, respectively, yields the same function.

Let v be an operator node of T_k that has variable x_i , $i \in [n]$, as child, and let α be the numerical value assigned to arc (v, x_i) . That is, $\alpha \cdot x_i$ is input of the operator associated with v . In G_k , arc (v, x_i) corresponds to the edge $\{v, v_i^k\}$ with color α . Due to the coloring of G_k and since π is an automorphism of G_k , there exists $j \in [\pm n]$ such that $\pi(v_i^k) = v_j^k$. Because of the same reasons, $\pi(v)$ corresponds to an operator of the same type as v and $\{\pi(v), v_j^k\}$ has color α . Thus, operator v in T_k has input $\alpha \cdot x_i$ and operator $\pi(v)$ in $\pi(T_k)$ has input

$$\alpha \cdot x_{\text{sgn}(i)|i|} = \text{sgn}(i)\alpha \cdot x_{|i|} = \alpha \cdot \gamma(x)_j,$$

where the first equality holds due to assigning edges $\{v, v_j^k\}$ and $\{v, v_{-j}^k\}$ negated values/colors. Consequently, since we assumed that all operators are commutative, assigning the pre-function of \bar{T}_k input x and the pre-function of $\pi(\bar{T}_k)$ input $\gamma(x)$ yields the same function. \square

Remark 4.4. Proposition 4.3 can easily be generalized to expression trees involving non-commutative operators. As mentioned in Section 3, the arcs leaving an operator node of such an expression tree need to be labeled to indicate the order in which the input of an operator needs to be processed. These labels then need to be incorporated into the edge colors of an SDG.

We close this section by evaluating the capabilities of the SDG of Proposition 4.3 for example (4). The reflection symmetries of (4) are the signed permutations $\gamma_1 = (1, -2)(2, -1)$, $\gamma_2 = (3, -4)(4, -3)$, $\gamma_3 = \gamma_2 \circ \gamma_1$, and $\gamma_4 = \text{id}$. The only signed permutations that can be derived from automorphisms of the SDG in Figure 5(a), however, are γ_1 and γ_4 . That is, even for linear expressions, the SDG might not allow to detect all reflection symmetries. This undesired behavior will be investigated in more detail in the next section, where we will modify the SDG to detect all reflection symmetries of linear expressions. Moreover, the modifications will also allow to detect the natural reflection symmetries of the disk packing problem from Example 1.1.

4.2 Enhancements of the Framework

Although SDGs are defined in terms of relative variable domains, the SDG of Proposition 4.3 for the exemplary MINLP (4) does not encode the reflection symmetry between x_3 and x_{-4} . The reason is that our proposed modifications of expression trees explicitly encode the original reflection centers in the expression trees. Variables with different reflection centers consequently cannot be symmetric in the SDG, cf. Figure 5(a). In this section, we discuss small modifications of the proposed SDGs, which allow to detect reflection symmetries of sum expressions, squared differences, bilinear products, and even operators. This list of examples is, of course, not exhaustive and only serves as an illustration of how the abstract concept of symmetry detection graphs can be used to derive tailored SDGs incorporating symmetry information of particular types of constraints.

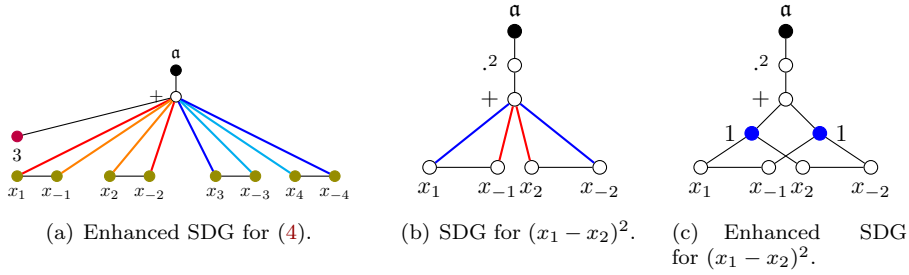


Figure 6: Illustration of a enhanced SDGs.

Sum Expressions. A simple way to resolve the issue for MINLP (4) is to not consider the summands of a sum expression independently. To make this precise, let $I \subseteq [n]$ and $S = \sum_{i \in I} \alpha_i \cdot x_i$, where $\alpha_i \in \mathbb{R}$ for all $i \in I$. To center each variable at the origin, the last section proposed to modify an expression tree by replacing each summand $\alpha_i \cdot x_i$ by a sub expression tree for $\alpha_i \cdot x_i + \alpha_i \cdot \xi_i$. Alternatively, we can compute the expression $S' = \sum_{i \in I} \alpha_i \cdot \xi_i + \sum_{i \in I} \alpha_i \cdot x_i$ first and create the expression tree for S' , see Figure 6(a). Since both the modified expression tree T from the previous section and the tree T' described here model the same function, automorphisms of the SDG corresponding to T' also correspond to reflection symmetries of (MINLP). In particular, because the SDG for S' essentially corresponds to the SDGs from [9] for linear constraints, all reflection symmetries of S are encoded in the SDG, see [9, Thm. 4].

Squared Differences. Consider Example 1.1 for two disks. Incorporating the previously discussed idea for sum expressions into SDGs allows to detect the symmetries $(x_1, x_2) \mapsto (-x_2, -x_1)$ and $(y_1, y_2) \mapsto (-y_2, -y_1)$ even when the box is not centered at the origin. The reflection along the horizontal and vertical symmetry axes of the box, however, cannot be detected via the SDG. The reason is that the SDG does not encode that $(x_1 - x_2)^2$ is the same as $(-x_1 - (-x_2))^2$ since the corresponding expression trees are different, see Figure 6(b) for an SDG of this expression.

To incorporate this information into an SDG, we replace the subgraph in the SDG corresponding to the sub expression tree for $(x_1 - x_2)^2$ by a gadget. For this gadget, we introduce a binary *squared difference* operator \boxminus modeling $\boxminus(x_1, x_2) = (x_1 - x_2)^2$. The gadget then consists of an operator node v of type \boxminus , two auxiliary value nodes a_1 and a_2 both receiving color 1, and the distinguished nodes v_1, v_{-1}, v_2, v_{-2} for x_1, x_{-1}, x_2, x_{-2} . The gadget's edges all remain uncolored and are given by $\{v, a_1\}$, $\{v, a_2\}$, $\{a_1, v_1\}$, $\{a_1, v_2\}$, $\{a_2, v_{-1}\}$, and $\{a_2, v_{-2}\}$, see Figure 6(c).

Since every SDG adds the edges $\{v_1, v_{-1}\}$ and $\{v_2, v_{-2}\}$, note that any automorphism of the gadget that exchanges v_1 and v_2 also needs to exchange v_{-1} and v_{-2} . Hence, the automorphism group of this gadget is generated (provided all variables have the same type) by the permutations $\pi_1 = (v_1, v_2)(v_{-1}, v_{-2})$ and $\pi_2 = (a_1, a_2)(v_1, v_{-1})(v_2, v_{-2})$. The former corresponds to the permutation symmetry of Example 1.1 that exchanges identical disks; the latter corresponds to the reflection along the vertical symmetry axis of the box (as we only consider x -variables). That is, all reflection symmetries of Example 1.1 can be detected via this gadget.

Note that the introduction of the new operator type indeed ensures that every automorphism of the SDG corresponds to a reflection symmetry of the corresponding MINLP, since the squared difference operator serves as an anchor of the gadget. In the following, we will generalize this idea to other structures frequently arising in MINLP. We only need to ensure that each newly introduced gadget is anchored and that there is a unique operator type identifying the gadget. The former ensures that the entire SDG remains anchored if a gadget replaces a sub expression tree, whereas the latter guarantees constraint compatibility among different SDGs.

Bilinear Products. Let $i, j \in [n]$ be distinct and consider the bilinear product $x_i \cdot x_j$. If the reflection center of both x_i and x_j is at the origin and both variables have the same type, the product admits two symmetries: either one exchanges x_i and x_j , or one simultaneously replaces x_i by x_{-i} and x_j by x_{-j} . These symmetries are exactly the same symmetries as for the case of

squared differences. For this reason, the symmetries can be encoded using an analogous gadget; the only difference is that the \boxminus operator node needs to be replaced by the product operator.

Even Functions. Let $i \in [n]$ and let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a univariate function such that $f(x_i)$ arises as a subexpression in (MINLP). If f is an even function, then $x_i \mapsto -x_i$ is a potential symmetry of (MINLP) as $f(x_i) = f(-x_i)$. This property of even functions can easily be incorporated into SDGs. Instead of connecting the operator node corresponding to f with the distinguished nodes v_i and v_{-i} by two edges being colored by the numerical values 1 and -1 , respectively, we assign both edges color 1.

The idea for even functions can be generalized, of course, if the input of f is a more complicated expression E . In this case, nodes v_i and v_{-i} must be replaced by the root nodes of graphs representing expression trees for E and $-E$, respectively. Depending on the size of E , this could increase the size of the SDGs significantly though, and thus also increases the time needed to detect symmetries.

5 An Open-Source Implementation of the Abstract Framework

In this section, we describe our implementation of the abstract symmetry detection framework of the previous section within the open-source solver SCIP. Our framework is contained as a C-implementation in the release of SCIP 9.0 [10] and replaces SCIP’s previous symmetry detection mechanism. We start by explaining the design principles of our implementation (Section 5.1), followed by an illustration of how to use our framework within SCIP (Section 5.2). The section is concluded in Section 5.3 by providing further technical details and a discussion of how to detect symmetries of SDGs. Appendix A provides an overview of the most important functions needed to apply our framework.

5.1 Design Principles

One of the design principles of SCIP is that all major components of the solver are organized as plug-ins. This allows users to easily extend SCIP by tailored techniques for specific applications, e.g., cutting planes or heuristics. The main plug-in type is a so-called constraint handler. A constraint handler provides an abstract notion of a class of constraints (e.g., general linear constraints, knapsack constraints, SOS1 constraints, or general nonlinear constraints), and defines general rules to enforce that a solution adheres to the corresponding constraints. That is, if SCIP solves an instance of (MINLP) and finds some intermediate solution x , every constraint g_k , $k \in [m]$, will ask its corresponding constraint handler for rules to check whether the solution satisfies $g_k(x) \leq 0$ and, if not, how this can be resolved.

With the release of SCIP 5.0, a symmetry detection mechanism had been implemented. This mechanism, however, was only able to detect symmetries if all constraints of a problem have a constraint handler being part of the SCIP release. One of our motivations to introduce the abstract notion of SDGs was to overcome this limitation and to allow symmetry detection also in the presence of custom constraints. We realized symmetry detection via SDGs in SCIP by introducing a new optional callback for constraint handlers. If a constraint handler implements this callback, the callback needs to create an SDG from the constraint’s data adhering to Proposition 4.3. The SDGs for individual constraints are then combined to a global SDG for the entire problem via Theorem 4.2. Symmetry detection in SCIP will then check whether the constraint handlers of all constraints present in the problem implement the new callback. If this is the case, symmetries are computed based on the global SDG; otherwise, symmetry computation is disabled. In the remainder of this section, we provide details of how SDGs can be encoded using our implementation, and how variable color and constraint compatibility can be ensured.

Nodes and Edges of SDGs. In the construction of SDGs for nonlinear constraints represented by expression trees, we used four different types of nodes: nodes representing numerical values, mathematical operators, variables, and an anchor. Since we believe that the representation via

expression trees is rather generic, our implementation also makes use of four different types of nodes that can hold different information:

value nodes store a floating-point number;

operator nodes store an integer value serving as an identifier of an operator;

variable nodes store the index of the corresponding (reflected) variable;

constraint nodes store a pointer to a constraint as well as two floating-point numbers (referred to as left-hand side and right-hand side).

Due to the definition of SDGs, our implementation automatically adds variable nodes for all (reflected) variables to an SDG. The remaining types of nodes can be added to an SDG by dedicated functions, see Appendix A. Each node of an SDG is identified by an integer index.

Edges $\{u, v\}$ of SDGs are identified by the indices u and v , and can optionally hold a floating-point value, cf. the assignment of numerical values to edges in Section 4.1. By calling a function, edges can be added to an SDG. Our implementation also makes sure that every SDG contains the edges connecting the distinguished nodes for a variable x_i and its reflection x_{-i} .

Note that neither for nodes nor for edges we allow to specify a color. The colors will be determined separately.

Ensuring Compatibility. To ensure constructing correct SDGs for optimization problems, all SDGs for individual constraints need to be anchored, and the collection of SDGs has to be variable color and constraint compatible. We therefore devised the following principles for constructing an SDG that should be used by symmetry detection callbacks of constraint handlers.

- (P1) The callback must not add edges connecting two variable nodes.
- (P2) Every SDG for an individual constraint should have exactly one constraint node that serves as an anchor.
- (P3) The SDGs constructed by a constraint handler for two constraints c, c' are only isomorphic if $c \equiv c'$, cf. Section 4.

Regarding (P1), the only edges between variable nodes that are required in Section 4 are the edges connecting variable nodes for pairs of reflected variables, which are present in an SDG by default in our implementation. Regarding (P2), although operator and value nodes could serve as anchors, too, constraint nodes easily allow to ensure constraint compatibility, see below.

A key component of ensuring compatibility is that only equivalent types of nodes and edges receive identical colors. We therefore did not allow to specify the color of a node or edge during creation, but instead assign properties to nodes and edges. Once all SDGs have been created, SCIP computes pairwise disjoint sets of colors for the edges and the four different types of nodes. While the colors of edges as well as operator and value nodes can be easily derived from the unique value assigned to the corresponding object, deriving the colors of variable and constraint nodes is not immediate. For constraint nodes, we derive a color based on the corresponding left-hand side and right-hand side values as well as the constraint handler of the associated constraint. Variable nodes are assigned a color based on their types as explained in Section 4. The latter guarantees variable color compatibility.

It remains to discuss constraint compatibility. Let G be the constructed SDG for an entire optimization problem, and let G' and G'' be the SDGs for two constraints of the problem. In the proof of Theorem 4.2, we noted that anchors guarantee that, if an automorphism of G maps a node of G' onto a node of G'' , then every node of G' needs to be mapped onto a node of G'' . Since the color of an anchor node depends on the corresponding constraint handler by (P2), only SDGs derived from the same constraint handler can be isomorphic. Principle (P3) then guarantees constraint compatibility.

Remark 5.1. Although (P3) seems difficult to achieve at first glance, it is usually easy to implement in practice. The reason is that constraint handlers need to implement abstract rules for deriving an SDG from a constraint. One such mechanism has been described for SDGs derived from expression trees in Section 4.1, and Section 5.2 will discuss another approach for particular constraints without an immediate expression tree representation.

5.2 Callbacks and Their Usage

In this section, we provide a more detailed description of our callbacks and we illustrate how to use them to detect symmetries.

Callbacks In SCIP 9.0, one can compute either reflection symmetries or classical permutation symmetries. To allow constraints to inform SCIP about how to detect (reflection) symmetries, their constraint handlers must implement the `SCIP_DECL_CONSGETSIGNEDPERMSYMGRAPH` callback for reflection symmetries or the `SCIP_DECL_CONSGETPERMSYMGRAPH` callback for permutation symmetries. Since the functionality of both callbacks is essentially the same, we refer to both of them just as “the callback” in the following. The only differences between the SDGs for permutation symmetries and reflection symmetries is that an SDG for permutation symmetries does not contain nodes for reflected variables and that the variable type is defined according to the original variable bounds (i.e., not shifted).

The input of either callback are five pointers providing the following information:

`scip` data structure with information about the problem and solving process;

`conshdlr` constraint handler for which the callback has been implemented;

`cons` constraint for which an SDG shall be created;

`graph` SDG to which the information of `cons` shall be added;

`success` Boolean to store whether the SDG could be created successfully.

If the callback cannot create the SDG, it should set `success` to `FALSE` to inform SCIP that not all constraints could provide symmetry information. Symmetry detection gets then disabled.

Constraints in SCIP usually provide further information that fully characterizes the constraint, so-called `consdata`. Linear constraints, for example, store the variables and coefficients as well as the left-hand side and right-hand side of the constraint in the `consdata`. The `consdata` can be accessed from `cons` and usually is the only information that should be used to create the SDG.

The SDG to which nodes and edges shall be added is given by `graph`. Note that we decided to not create a separate SDG for every constraint. Instead, `graph` corresponds to the SDG for the entire problem. This does not cause any conflicts as long as the callback only adds edges between nodes that have been created by the callback for the same constraint. Since no new variable nodes can be created, there cannot be multiple nodes modeling the same variable.

Using the Callback We illustrate how the callbacks can be used for the stable set problem. Let $H = (V, E)$ be an undirected graph with node weights $w_v \in \mathbb{R}$, $v \in V$. A *stable set* in H is a set $S \subseteq V$ if $\{u, v\} \notin E$ for all $u, v \in S$. The task is to find a stable set S with maximum total node weight. A classical integer programming formulation is

$$\max \left\{ \sum_{v \in V} w_v x_v : x_u + x_v \leq 1 \text{ for all } \{u, v\} \in E \text{ and } x \in \{0, 1\}^V \right\}.$$

It is well-known that the linear programming (LP) relaxation of this formulation is rather weak and that it can be enhanced by adding so-called clique inequalities $\sum_{v \in C} x_v \leq 1$, where C is a clique in H . Among others, the following two possibilities exist to make use of clique inequalities in SCIP. On the one hand, one could enumerate all clique inequalities explicitly and add them as linear inequalities to the problem formulation. In this case, SCIP can automatically detect symmetries. But since there might be exponentially many cliques, the LP relaxation could become much harder to solve. On the other hand, one could implement an abstract constraint handler that gets H as input and decides on the fly whether a solution satisfies all clique inequalities, and adds an inequality in case it is violated by the solution. Without our callback, however, SCIP cannot know how the presence of the stable set constraint handler impacts the symmetries of the problem.

To implement our callback, we observe that every automorphism of H that exchanges nodes of the same weight is also a symmetry of the stable set problem. The SDG for the stable set

```

1  static
2  SCIP_DECL_CONSGETPERMSYMGGRAPH(consGetPermsymGraphStableSet)
3  {
4      SCIP_CONSDATA* consdata;
5      int* idx;
6      int vidx;
7      int nnodes;
8      int nodeop;
9      int v;
10
11     /* define node operator */
12     nodeop = 0;
13
14     consdata = SCIPconsGetData(cons);
15     nnodes = consdata->nnodes;
16
17     SCIP_CALL( SCIPallocBufferArray(scip, &idx, nnodes + 1) );
18
19     /* create operator nodes and constraint node */
20     for( v = 0; v < nnodes; ++v )
21     {
22         SCIP_CALL( SCIPaddSymgraphOpnode(scip, graph, nodeop, &idx[v]) );
23     }
24     SCIP_CALL( SCIPaddSymgraphConsnode(scip, graph, cons, 0.0, 0.0, &idx[nnodes]) );
25
26     /* add edges of underlying graph */
27     for( v = 0; v < consdata->nedges; ++v )
28     {
29         SCIP_CALL( SCIPaddSymgraphEdge(scip, graph,
30             idx[consdata->first[v]], idx[consdata->second[v]], FALSE, 0.0) );
31     }
32
33     /* connect nodes with constraint node */
34     for( v = 0; v < nnodes; ++v )
35     {
36         SCIP_CALL( SCIPaddSymgraphEdge(scip, graph,
37             idx[v], nodeidx[nnodes], FALSE, 0.0) );
38     }
39
40     /* connect operator nodes with variable nodes, assign edges weight of node */
41     for( v = 0; v < nnodes; ++v )
42     {
43         vidx = SCIPgetSymgraphVarnodeidx(scip, graph, consdata->vars[v]);
44         SCIP_CALL( SCIPaddSymgraphEdge(scip, graph,
45             idx[v], vidx, TRUE, consdata->weights[v]) );
46     }
47     *success = TRUE;
48
49     SCIPfreeBufferArray(scip, &idx);
50
51     return SCIP_OKAY;
52 }

```

Figure 7: Illustration of usage of symmetry detection callback.

constraint handler thus essentially needs to add a copy of H to the global SDG. We illustrate this in Figure 7 and provide a description next.

We assume that the consdata of the stable set constraint contains the following information: **nnodes** provides the number of nodes in H and the nodes are labeled $0, \dots, \text{nnodes} - 1$; **nedges** provides the number of edges in H and edges are encoded via two arrays **first** and **second** that contain the first and second nodes of all edges, respectively; **weights** is an array that assigns each node its weight.

First, we encode an operator type in Line 12 of Figure 7, which represent nodes of the graph H . Note that there might exist other constraint handlers that also define an operator with the same index. This is not an issue though as (P2) will make sure that we cannot mix constraints of the stable set constraint handler with constraints from other constraint handlers.

Second, we extract some information about H and create an array `idx` to store the indices of the newly created nodes of the SDG (Line 17). Afterwards, we create for each node of H a copy in the SDG by creating a corresponding operator node (Line 22). We also create an anchor of the SDG (Line 24) by introducing a constraint node. Since the abstract stable set constraint has neither a left-hand side nor a right-hand side, we store the dummy values 0 and only assign the constraint pointer `cons` to the anchor node.

Third, the edges of H are copied to the SDG in Line 29. Since edges of H are unweighted, also the corresponding edges of the SDG are unweighted (indicated by `FALSE`). Fourth, we connect the anchor with the remaining nodes of the SDG. To preserve symmetries of H , the anchor is connected with all copied nodes of H (Line 36). Finally, we add edges between the copies of nodes of H and the corresponding variable nodes (Line 44). These edges receive the weights of the corresponding nodes in H , and we set the `success` pointer to `TRUE` since the graph could be created.

5.3 Technical Details

In this section, we provide further technical details of our implementation. First, we briefly discuss our data structures for SDGs. Second, we mention further details of the implementation to realize the ideas of Section 4 in SCIP. Finally, we explain how automorphism of SDGs can be detected.

Encoding Symmetry Detection Graphs We experimented with different data structures for encoding SDGs. Our first attempt was to use an object-oriented implementation containing separate objects (or structs in `C`) for graphs, nodes, and edges. But it turned out that, even for linear problems, the creation of SDGs was much slower than the previous implementation in SCIP that was tailored towards constraints known by SCIP. The main reason for the slower performance was an overhead in memory allocation for the separate objects. For the same reason, we discarded the idea of creating separate SDGs for individual constraints. Instead, we only maintain a single SDG that is extended by the callbacks, which turned out to be more efficient. The final data structure for SDGs is a `C` struct using basic `C` data types such as `int`, `double`, or pointers to encode information about nodes and edges.

In the following, assume an SDG has n nodes and m edges. Nodes of SDGs are identified by an index in $N = \{0, \dots, n - 1\}$ and several arrays store information about the nodes. Recall that we distinguish four different types of nodes (operator, numerical value, variable, and constraint). These types are encoded by an `enum` and several arrays hold the information corresponding to the nodes of different types. For example, array `vals` holds the values assigned to value nodes, and array `lhs` stores the left-hand side values assigned to the constraint nodes. To access the information associated with node $i \in N$, we use two arrays. Array `nodetypes` stores the type of the different nodes and array `nodeinfopos` stores the index of each node in its corresponding group. That is, if j is the value stored at `nodeinfopos[i]`, then node $i \in N$ is the j -th node within the group `nodetypes[i]`. If node i was a constraint node, we thus could access its associated left-hand side at `lhs[j]`. Several functions are available to access node information or to create nodes, see Appendix A. When creating a new node, we check whether the node still fits into the allocated memory for the aforementioned arrays. If not, we enlarge the arrays by memory reallocation; the new size of the arrays is determined by SCIP based on a growth factor to avoid frequent memory reallocation.

Edges are identified by an index in $\{0, \dots, m - 1\}$. Two arrays `edgefirst` and `edgesecond` store the indices of the first and second nodes in the edges, respectively; array `edgevals` stores the values associated with an edge. If an edge has no assigned value, we use the placeholder value infinity in `edgevals`.

Moreover, an SDG struct contains `int` arrays for the colors associated with nodes and edges. Initially, these arrays are empty and only created if computing colors is triggered by calling a function, see the next section for details. In this case, an SDG will lock itself, which means that it is no longer possible to add nodes and edges. This is a safety mechanism to ensure that the stored colors always correspond to the SDG.

Implementation Details We discuss two implementation details concerning numerical inaccuracies and aggregation of variables. The former is relevant for computing colors, whereas the

second has consequences for creating SDGs.

To derive colors for nodes, we sort nodes first by their type (operator, constraint etc.) and then based on the information associated with them. For instance, numerical value nodes are sorted based on their associated value, and variable nodes are sorted based on their variable’s type. Since SCIP uses floating-point arithmetic when solving an optimization problem, it is reasonable to also consider two nodes as identical if their associated information does not deviate more than a small quantity ε (by default 10^{-9} in SCIP). We then partition the sorted list of nodes into blocks of consecutive elements such that the first and last node per block do not deviate more than ε . Every node within the same block is then assigned the same color, and nodes from different blocks receive different colors. Colors for edges are computed analogously.

In Definition 4.1, we required SDGs to contain a distinguished node for every (reflected) variable of (MINLP). In practice, however, SCIP removes variables from a problem, e.g., when they get fixed or can be represented as a weighted sum of other variables (variable aggregation). To reduce the number of variable nodes in SDGs, it thus makes sense to only represent variables in an SDG that are neither fixed nor aggregated. Our implementation of SDGs therefore also only supports unfixed and non-aggregated variables. When using the symmetry detection callbacks, every occurrence of a fixed or aggregated variable needs to be replaced by the corresponding constant or sum of variables, respectively.

Detecting Automorphisms To detect symmetries of SDGs, we use external software packages for computing graph automorphisms. Our current implementation supports the packages BLISS [22] and NAUTY [36]. Since BLISS and NAUTY only support graphs with uncolored edges, we post-process SDGs: every edge $\{u, v\}$ of color c is replaced by an auxiliary node w of color c as well as the uncolored edges $\{u, w\}$ and $\{v, w\}$. As noted in Section 3, the number of auxiliary nodes can be reduced by identifying some of them with each other. For the SDGs G for permutation symmetries of MILP, see Section 3, this can be achieved as follows, cf. [41].

Let v be a variable node of G and let E_c^v be the set of all edges of color c that are incident with v . Instead of introducing an auxiliary node for all edges in E_c^v , it is sufficient to introduce one auxiliary node w as well as the edges $\{v, w\}$ and $\{u, w\}$ for all $\{u, v\} \in E_c^v$. This way, the number of auxiliary nodes changes from $|E_c^v|$ to 1 and the number of edges changes from $2 \cdot |E_c^v|$ to $1 + |E_c^v|$. This mechanism of reducing the number of auxiliary nodes is called *grouping by constraints* [40]; if the mechanism is applied to constraint nodes instead of variable nodes, it is called *grouping by variables*.

We have implemented an analogous grouping mechanism for our abstract SDGs, where we group edges being incident to either the same constraint or variable node. The former is used if there are less constraint nodes than variable nodes.

6 Handling Reflection Symmetries

In this section, we describe methods for handling (reflection) symmetries in MINLP, which we illustrate using the disk packing problem of Example 1.1. In the disk packing problem, we distinguish three types of symmetries: (i) permutation symmetries that exchange indices of disks; (ii) reflection symmetries that reflect the x - or y -coordinates of all disks; (iii) and, if the width and height of the box are identical, permutation symmetries that exchange both coordinates. To visualize these symmetries, it is convenient to encode a solution of the disk packing problem by a matrix $M \in \mathbb{R}^{D \times 2}$, where the i -th row (M_{i1}, M_{i2}) corresponds to the center point (x_i, y_i) of the i -th disk. The first class of symmetries then corresponds to reordering the rows of M , the second class to changing the signs of all entries of a column of M , and the last class to an exchange of the two columns of M , see Figure 8.

The described actions of the symmetry group of an MINLP on a matrix of variables is rather generic and, as we will see in Section 7, occurs in many applications. We therefore discuss, after a short overview of symmetry handling techniques for permutation symmetries, methods for row-exchanges, column-reflections, and their combination with column-exchanges in more detail. We conclude by devising general techniques for handling reflection symmetries.

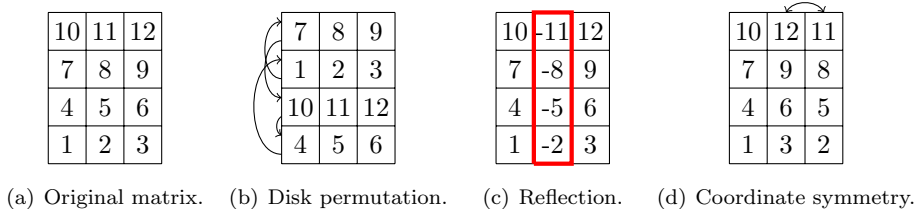


Figure 8: Illustration of matrix symmetries arising in Example 1.1.

6.1 General Symmetries

Consider an (MINLP) with variable vector $x \in \mathbb{R}^n$ and reflection symmetry group $\Gamma \leq \mathcal{S}_n^\pm$. A standard way for handling Γ is to define an order of the variables, encoded by a permutation $\pi \in \mathcal{S}_n$, and to enforce that a solution x must be lexicographically maximal in its Γ -orbit. Formally, one is looking for feasible solutions x of (MINLP) satisfying $\pi(x) \geq_{\text{lex}} \pi(\gamma(x))$ for all $\gamma \in \Gamma$, where \geq_{lex} denotes the lexicographic comparison. In general, enforcing the lexicographic order constraints in coNP-hard, cf. [3]. One therefore typically uses one of the following three alternatives.

First, one can handle the lexicographic order constraints for a small subset of symmetries. For binary variables, a constraint $\pi(x) \geq_{\text{lex}} \pi(\gamma(x))$ can be handled by adding linear inequalities to (MINLP). To fully encode the lexicographic comparison by linear inequalities, however, one needs exponentially many inequalities or inequalities with exponentially large coefficients [20]. An exponentially large class with coefficients in $\{0, \pm 1\}$ is presented in [21], whereas [16] uses a single inequality with exponentially large coefficients. To overcome these drawbacks, an alternative is to propagate lexicographic constraints during branch-and-bound. That is, given the bounds on variables of a subproblem of branch-and-bound, one derives further reductions of variable bounds that are satisfied by any lexicographic maximal solution. Such a propagation algorithm, *lexicographic reduction*, is described in [13] and runs in linear time. In particular, the algorithm also works for non-binary variables.

Second, one can derive strong symmetry handling methods for particular classes of symmetry groups. Besides techniques for the row symmetries of the disk packing problem, which we discuss in the next section, also efficient propagation algorithms for particular cyclic groups exist [14].

Third, there exist intermediate approaches that handle some but not necessarily all group structure. For example, *orbital fixing* [39] fixes binary variables based on the branching decisions and permutation symmetries of (MINLP). Recently, this method has been generalized to arbitrary variable types [13], so-called *orbital reduction*. Moreover, [31, 45] discuss sparse linear inequalities that are derived from the Schreier-Sims table of a permutation group and that handle symmetries based on variable orbits.

Recall that, by Definition 2.2, symmetric variables can have different variable domains due to translations. For handling row and column symmetries, we assume in the following that all symmetric variables have the same domain.

6.2 Row Symmetries

A popular approach for handling row symmetries of a matrix $M \in \mathbb{R}^{p \times q}$ is to lexicographically sort its rows, i.e., if M_i , $i \in [p]$, denotes the i -th row of M , one enforces $M_1. \geq_{\text{lex}} M_2. \geq_{\text{lex}} \dots \geq_{\text{lex}} M_p.$. By interpreting M as a vector $x \in \mathbb{R}^{pq}$ such that entry (i, j) of M is entry $(i-1)q + j$ of x , this can be achieved via the lexicographic ordering constraints of the previous section using $\pi = \text{id}$. But since the group of all row permutations has size $p!$, it is impractical to add all ordering constraints. Instead, it is folklore that sorting the rows can already be achieved by the ordering constraints for the $p-1$ permutations that exchange consecutive rows, cf. [21]. Methods such as lexicographic reduction can then be used to enforce the ordering constraints.

Since these methods ignore the interplay of different row permutations, they might not detect all reductions that can be derived from a matrix with sorted rows. An alternative is *orbitopal reduction* [13]. This propagation algorithm is called within a branch-and-bound algorithm and receives bounds on all variables in M w.r.t. the current subproblem. The algorithm then identifies

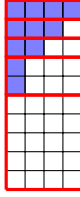


Figure 9: Row symmetries combined with reflection symmetries of columns. Rows within the red blocks can be exchanged arbitrarily, whereas blue cells can be restricted to the upper half of the domain.

the tightest bounds for all variables in M to which every sorted feasible solution of the subproblem adheres to. Originally, this algorithm has been derived for matrices of binary variables [5], but the variant of [13] is applicable for general integer and continuous variables, too. Moreover, [24] describe a variant of this algorithm if all variables in M are binary and at most one (resp. exactly one) variable per column is allowed to attain value 1. A facet description of the convex hull of all such binary matrices M is also known [25]; the facet-defining inequalities can then be added to (MINLP) to handle row symmetries.

Above, we discussed a fixed assignment of the entries of $M \in \mathbb{R}^{p \times q}$ to a vector $x \in \mathbb{R}^{pq}$. This requirement can be relaxed such that different assignments can be used at different nodes of the branch-and-bound tree. For example, [5] discusses a mechanism to change the order of the columns of M before translating M into a vector x . A variant in which also the order of the rows can be exchanged is discussed in [13]. We refer to both variants as *column-dynamic* and *column-row-dynamic* orbitopal fixing, respectively. The variant without any reordering is referred to as *static* orbitopal fixing.

6.3 Column Reflections

Let $M \in \mathbb{R}^{D \times 2}$ be the variable matrix for the disk packing problem for D disks. Then, M admits row symmetries and reflection symmetries of individual columns. Let $n_1 = \lceil \frac{D}{2} \rceil$ and $n_2 = \lfloor \frac{D}{2} \rfloor$. To handle reflection symmetries in the regime of disk packing, [26] enforces $M_{i,1} \geq 0$ for all $i \in [n_1]$ and $M_{i,2} \geq 0$ for all $i \in [n_2]$. Geometrically, this means that there are at least as many disks in the right half of the box than in the left half; moreover, within the right half, there are not less disks in the upper half than in the lower half.

Formally, we can derive these reductions as follows. After possibly applying a reflection of the first column, one can guarantee that the first column has at least as many non-negative entries as non-positive entries. Due to row symmetries, the first n_1 entries can thus be enforced to be non-negative. This argument can be repeated for the first n_1 rows of the second column, because enforcing the non-negativity structure on the first column does not allow to exchange one of the first n_1 rows of M with one of the last $D - n_1$ rows.

The latter argument also shows that one can sort the first n_1 rows of M and the last $D - n_1$ rows lexicographically. To partially enforce this, [26] suggests to add the inequalities $M_{i,1} \geq M_{i+1,1}$ for $i \in [D - 1] \setminus \{n_1\}$. A similar sorting idea has been applied for the kissing number problem in [29].

The idea of [26] can be generalized to matrices with more than two columns. We provide the following result without a proof, see Figure 9 for an illustration.

Proposition 6.1. *Let $M \in \mathbb{R}^{p \times q}$ be a variable matrix admitting row symmetries and reflection symmetries for individual columns. Suppose that, for all $j \in [q]$, all variables in column j of M have the same domain center $\xi_j \in \mathbb{R}$. Let $n_0 = p$ and, for $j \in [q]$, let $n_j = \lceil \frac{n_{j-1}}{2} \rceil$. Then, a valid symmetry handling approach is given by*

- adding, for each $j \in [q]$ and $i \in [n_j]$, the inequality $M_{i,j} \geq \xi_j$, and
- enforcing, for each $j \in [q] \cup \{0\}$, that $M_{n_j+1,\cdot} \geq_{\text{lex}} \dots \geq_{\text{lex}} M_{n_{j-1},\cdot}$.

The lexicographic sorting can be enforced, e.g., via static orbitopal reduction. A dynamic variant of orbitopal reduction is not immediately applicable due to the additional lower bound constraints derived from the reflection symmetries.

6.4 Row and Column Symmetries

In contrast to row symmetries only, it is coNP-complete to decide whether a matrix $M \in \{0, 1\}^{p \times q}$ is lexicographically maximal w.r.t. row and column symmetries [8]. There is consequently no efficient mechanism to simultaneously handle row and column symmetries unless $P = \text{coNP}$. A possible remedy to handle at least some symmetries is to enforce that both the rows and columns of a solution matrix $M \in \mathbb{R}^{p \times q}$ of (MINLP) are sorted lexicographically [15].

Sorting the rows can be achieved by static orbitopal reduction as described before. Moreover, sorting of the columns can be achieved by applying static orbitopal reduction to the transposed variable matrix. This is indeed compatible as one checks via the lexicographic ordering constraints: Recall that we associated with $M \in \mathbb{R}^{p \times q}$ the vector $x \in \mathbb{R}^{pq}$ such that $M_{i,j}$ corresponds to $x_{(i-1)p+j}$. As mentioned above, the rows of M are sorted lexicographically if and only if $x \geq_{\text{lex}} \gamma(x)$ for all row permutations γ . However, one can show that also the vector x' that identifies $M_{i,j}$ with $x'_{(j-1)q+i}$ has the same properties. That is, since x' corresponds to the first mechanism for the transpose of M , orbitopal reduction for both M and its transpose can be combined.

Due to the interaction of row and column symmetries, we stress that a dynamic variant of orbitopal reduction is not immediately applicable. Moreover, if also reflection symmetries act on individual columns, handling row symmetries can be replaced by the mechanism in Proposition 6.1.

6.5 Further Techniques for Handling Reflection Symmetries

We describe two further techniques for handling reflection symmetries that also apply if no row or column symmetries are present in the problem. The first one is a simple inequality that handles some, but by far not all symmetries. Consider an instance of (MINLP) that contains the signed permutation γ^* that reflects *all* variables simultaneously, i.e., $\gamma^*(i) = -i$ for all $i \in [n]$. Instead of lexicographically comparing a solution x of (MINLP) with $\gamma^*(x)$, one can enforce that the aggregated weight of x is larger than the weight of $\gamma^*(x)$, i.e.,

$$\sum_{i=1}^n x_i \geq \sum_{i=1}^n \gamma^*(x)_i = \sum_{i=1}^n (2\xi_i - x_i) \quad \Leftrightarrow \quad \sum_{i=1}^n x_i \geq \sum_{i=1}^n \xi_i. \quad (5)$$

This inequality can, e.g., be used for the max-cut problem [27], where the aim is to partition the node set of an undirected graph $G = (V, E)$ into two sets such that the weight of edges between both sets is maximized. If a binary variable x_v , $v \in V$, indicates whether v is contained in the first set ($x_v = 1$) or not ($x_v = 0$), all variables are symmetric w.r.t. the described reflection symmetry. Inequality (5) then models that at least as many nodes are in the first set of the partition than in the second set.

Note that (5) cannot necessarily be combined with lexicographic order based techniques. For example, vector $e^1 = (1, 0, \dots, 0)$ satisfies $e^1 \geq_{\text{lex}} \gamma^*(e^1)$, but violates (5). We therefore also discuss a second approach: a generalization of lexicographic reduction to reflection symmetries.

The lexicographic reduction algorithm of [13] receives a permutation symmetry π of (MINLP) as well as upper and lower bounds for all variables as input. The algorithm then finds, for each variable, the tightest lower and upper bounds such that any solution x satisfying the initial bounds and $x \geq_{\text{lex}} \pi(x)$ adheres to the new bounds. Their algorithm can be immediately generalized to reflection symmetries $\gamma \in \mathcal{S}_n^\pm$. Since the arguments for correctness are the same as in [13], we only provide the high-level ideas but no formal proof.

The core of lexicographic reduction is the following observation: Suppose there is $i \in [n]$ such that $x_j = \gamma(x)_j$ for all $j \in [i-1]$. Then, $x \geq_{\text{lex}} \gamma(x)$ can only hold if $x_i \geq \gamma(x)_i$. In this case, the lower bound of variable x_i can possibly be strengthened to the lower bound of $\gamma(x)_i$ and, vice versa, the upper bound of $\gamma(x)_i$ can possibly be strengthened to the upper bound of x_i .

To exploit this observation for $\gamma \in \mathcal{S}_n^\pm$, lexicographic reduction iterates over the entries x_i , $i \in [n]$, of a solution x and checks whether $x_j = \gamma(x)_j$ holds for all $j \in [i-1]$ by comparing the

upper and lower bounds of the variables. If this is the case, three cases are distinguished. First, the variable bounds imply $x_i < \gamma(x)_i$. Then, the algorithm reports infeasibility because the variable bounds imply $x \not\geq_{\text{lex}} \gamma(x)$. Second, the bounds imply $x_i > \gamma(x)_i$. Then the algorithm stops since every solution adhering to the bounds satisfies $x >_{\text{lex}} \gamma(x)$. Third, the upper and lower bounds on the variables can possibly be improved due to the observation and the algorithm continues with the next iteration and strengthened bounds.

If the algorithm terminates and did not report infeasibility, lexicographic reduction has possibly improved some variable bounds. As discussed in [13], these bounds are as tight as possible for all variables except for the lower bound on x_i and the upper bound on $\gamma(x)_i$ of the last iteration i . In a post-processing step, lexicographic reduction can determine whether these bounds can be improved further, see [13] for details. The running time of lexicographic reduction is $O(n)$.

7 Numerical Experience

In this section, we evaluate our symmetry detection framework when solving MILPs and MINLPs. Specifically, we aim to answer the following questions:

- (Q1) Does the framework of Sections 4 and 5 reliably detect reflection symmetries?
- (Q2) What is the most effective way to enforce the lexicographic order constraints in Proposition 6.1?
- (Q3) How frequently arise reflection symmetries in benchmarking instances?
- (Q4) Is there a computational benefit for benchmarking instances when handling reflection symmetries instead of permutation symmetries?

To answer (Q1), we apply our framework to specific applications in which we know that reflection symmetries arise and we evaluate whether these symmetries can be detected (Section 7.2). Many of these instances contain row and column symmetries, which will allow us to answer (Q2). In Section 7.3, we turn the focus to instances from three benchmarking test sets, which allows us to answer (Q3) and (Q4). Before answering the questions, we provide an overview of how symmetries are detected and handled in Section 7.1.

7.1 Computational Setup

As mentioned in Section 5, we included a C-implementation of our detection framework into the solver SCIP. Our implementation is publicly available in SCIP; detailed instructions on how to reproduce our results are available at [18]. For most of SCIP's default constraint handlers, we have implemented the two callbacks for detecting permutation and reflection symmetries. SCIP allows to compute symmetries of the original problem or of the problem after presolving. In our experiments, we considered symmetries of the presolved problem, which is also SCIP's default setting. SCIP then builds the corresponding SDG, computes its automorphism group, and returns a set Γ' of signed permutations (so-called *generators*) that generates a group Γ of symmetries of the MINLP. That is, we do not have explicit access to all symmetries in Γ , but we can write every $\gamma \in \Gamma$ as a finite composition of elements from Γ' .

SCIP offers a variety of state-of-the-art methods for handling symmetries. To decide which method is used, SCIP analyzes the symmetry group $\Gamma \leq \mathcal{S}_n^\pm$ and checks whether it is the direct product of smaller groups, i.e., $\Gamma = \bigotimes_{s=1}^t \Gamma_s$ for some signed permutation groups Γ_s , $s \in [t]$. Each factor Γ_s can then be handled independently from the others, and SCIP calls different heuristics for each factor to decide how the symmetries are handled. The heuristics aim to detect structured symmetry groups in the following order:

1. row and column symmetries;
2. row symmetries;
3. unclassified symmetry groups.

We briefly describe how we detect and handle these symmetries. For the ease of notation, we assume that Γ has a single factor.

Row and Column Symmetries. The group of row symmetries is generated by permutations that exchange pairs of consecutive rows of a matrix, cf. Section 6.2. If $x, y \in \mathbb{R}^q$ denote such rows, the corresponding row permutation is $(x_1, y_1)(x_2, y_2) \dots (x_q, y_q)$, i.e., it is a composition of cycles of length 2. Analogously, the group of column symmetries is generated by exchanges of consecutive columns. We refer to the number of 2-cycles as the composition’s length.

When detecting row and column symmetries, however, the underlying matrix is unknown. Our heuristic therefore collects all generators of Γ corresponding to permutation symmetries. If there is a permutation that is not a composition of 2-cycles, the heuristic stops: no row/column symmetries are detected. Otherwise, the permutations are partitioned based on the length of their composition. If this partition consists of two sets, the heuristic tries to construct a matrix $M \in \mathbb{R}^{p \times q}$ such that the permutations of the first and second set correspond to row and column symmetries of the matrix, respectively.

If row and column symmetries are detected, one can handle symmetries by enforcing that the rows and columns of the matrix are sorted lexicographically, cf. Section 6. Moreover, the heuristic also checks whether there is a permutation that corresponds to the reflection of a single column (or row). Since all columns (or rows) are symmetric, this means that the group Γ also contains the remaining column (or row) reflections. In this case, we can strengthen the symmetry handling approach by restricting the variable domain of some variables according to Proposition 6.1.

To enforce sorted rows and columns in M in the absence of reflections, we explicitly add the inequalities $M_{i,1} \geq M_{i+1,1}$ for all $i \in [p-1]$ and $M_{1,j} \geq M_{1,j+1}$ for all $j \in [q-1]$, respectively, to the MINLP. Moreover, we use static orbitopal reduction to sort the rows and columns. Note that the inequalities are in principle also enforced by orbitopal reduction. Preliminary experiments showed, however, that the dual bounds can improve when adding the inequalities explicitly. This is in one line with an observation by [12] that symmetry handling inequalities can strengthen relaxations for MINLPs. In the presence of reflection symmetries for columns, we also add inequalities and use orbitopal reduction to sort the rows, but we restrict both techniques to the blocks of symmetric rows as indicated by Proposition 6.1. Additionally, we restrict some variable domains to their upper half as described in Proposition 6.1. If there are row reflections instead of column reflections, we proceed analogously.

Row Symmetries. To detect row symmetries, we use an analogous mechanism as for row and column symmetries. The only difference is that the heuristic can only be successful if all permutations have the same number of 2-cycles.

In the presence of column reflections, row symmetries are handled analogously to the case of row and column symmetries. If no column reflections are detected, we only handle row symmetries if the number of generators in Γ' being permutation symmetries is greater than 80%. Our reasoning is that if the percentage of permutation symmetries is too low, we potentially miss many symmetry reductions being implied by proper reflection symmetries. In this case, we proceed with general techniques as explained in the next paragraph.

If row symmetries for a matrix $M \in \mathbb{R}^{p \times q}$ are handled, we use dynamic orbitopal fixing besides the following exceptions:

- If there is only one column, we add the inequalities $M_{1,1} \geq \dots \geq M_{p,1}$ to fully handle the symmetries.
- If there are exactly two rows, we use lexicographic reduction to handle the symmetries of the corresponding permutation exchanging the two rows.
- Suppose there are at least three rows, and there are at least three columns containing only binary variables such that other problem constraints enforce that the column sums are exactly or at most 1. Then, we use orbitopal fixing for packing and partitioning orbitopes [24] for the submatrix containing the cardinality restricted columns.

These are also the default settings in SCIP to handle row symmetries. The motivation for the first, second, and third point is to add strong symmetry handling inequalities, more flexible symmetry handling techniques, or techniques exploiting additional problem structure, respectively. The hope is that these techniques are more powerful than dynamic orbitopal reduction.

General Symmetries. We test two different approaches to handle reflection symmetries. The first approach uses orbital reduction and lexicographic reduction; the second approach just adds (5) if it is applicable and no other symmetry handling method is used.

To apply the first approach, let Π' be the generators in Γ' that are permutation symmetries. The group generated by Π' is then handled by orbital reduction as implemented in SCIP. Moreover, for each generator $\gamma \in \Gamma'$, we enforce a lexicographic order constraint $\pi(x) \geq_{\text{lex}} \gamma(\pi(x))$ by lexicographic reduction, cf. Section 6.5. The reordering π can differ at each node of the branch-and-bound tree and is defined based on the branching history of the respective node. This allows to apply lexicographic reduction and orbital reduction simultaneously [13].

Hardware and Software Specifications. All of the following experiments have been conducted on a Linux cluster with Intel Xeon E5-1620 v4 3.5 GHz quad core processors and 32 GB memory. The code was executed using a single thread and the time limit for all computations was 2 h per instance. A memory limit of 27 GB has been used per instance.

We use a developers version of the branch-and-bound framework SCIP 10.0 (githash a93d088d). All LP relaxations are solved with `Soplex` 8.0 (githash 89ab43a) and nonlinear problems are solved with `Ipopt` 3.12 [48]. To detect symmetries of SDGs, we use the software `sassy` 1.1 [2] to preprocess SDGs; symmetries of the preprocessed SDGs are computed using `bliss` 0.77 [23].

To compare different settings in our experiments, we use, among others, mean numbers. All mean numbers for quantities t_1, \dots, t_n are reported in shifted geometric mean $\prod_{i=1}^n (t_i + s)^{\frac{1}{n}} - s$ to reduce the impact of outliers. For mean running times, a shift of $s = 1$ is used; otherwise, we use a shift of $s = 0$, i.e., the classical geometric mean.

7.2 Results for Structured Instances

To answer Question (Q1), we consider four different classes of instances with different types of nonlinearities that all admit reflection symmetries. Moreover, three of them admit row and column symmetries. The four considered problems, which we discuss next, are a geometric packing problem, the kissing number problem, an energy minimization problem, and the max-cut problem.

Geometric Packing Problem. Let d and n be positive integer numbers. We aim to find the largest real number r such that n non-overlapping d -dimensional ℓ_1 -balls of radius r can be packed into $[-1, 1]^d$. This problem can be modeled as

$$\begin{aligned} \max r \\ \sum_{i=1}^d |x_i^s - x_i^t| &\geq 2r, & s, t \in [n], s < t, \\ -1 + r \leq x_i^s &\leq 1 - r, & s \in [n], i \in [d], \end{aligned}$$

where r models the radius of the balls and x_i^s , $(i, s) \in [d] \times [n]$, is the i -th coordinate of the s -th ball's center. Interpreting x_i^s as the entries of an $n \times d$ matrix, this formulation admits row and column symmetries as well as reflection symmetries of individual columns. Although this model can easily be linearized, we do not do so to check whether symmetry detection works for constraints involving absolute values.

Kissing Number Problem. Let d and n be positive integers and let S be the unit sphere in \mathbb{R}^d w.r.t. the ℓ_2 -norm that is centered at the origin. The kissing number problem asks whether n unit spheres can be placed in \mathbb{R}^n such that each of the n spheres touches S and neither pair of the n spheres intersects. An optimization variant of this problem is looking for the maximum distance

of the center points of the n spheres [29]:

$$\begin{aligned}
& \max \alpha \\
& \sum_{i=1}^d (x_i^s)^2 = 4, & s \in [n], \\
& 8 - 2 \sum_{i=1}^d x_i^s x_i^t \geq 4\alpha, & s, t \in [n], s < t, \\
& -2 \leq x_i^s \leq 2, & s \in [n], i \in [d], \\
& 0 \leq \alpha \leq 1,
\end{aligned}$$

where the interpretation of x_i^s is the same as for the geometric packing problem. The second constraint indeed measures the distance between two center points as $\|x^s - x^t\|_2^2 = 8 - 2 \sum_{i=1}^d x_i^s x_i^t$. As before, this problem admits row and column symmetries with column reflections.

Energy Minimization Problem. In the energy minimization problem, we consider n particles that need to be distributed on a d -dimensional unit sphere [42]. If x_i^s models the i -th coordinate of the s -th particle, the energy $\frac{1}{\sum_{i=1}^d (x_i^s - x_i^t)^2}$ is needed to keep particles $s, t \in [n]$ at their distance. The goal is to distribute the particles such that their total energy is minimized, i.e., to solve

$$\begin{aligned}
\min \sum_{s=1}^n \sum_{t=s+1}^n \frac{1}{\sum_{i=1}^d (x_i^s - x_i^t)^2} \\
& \sum_{i=1}^d (x_i^s)^2 = 1, & s \in [n], \\
& -1 \leq x_i^s \leq 1, & s \in [n], i \in [d].
\end{aligned}$$

Also this problem admits row and column symmetries with column reflections.

Max-Cut. Given an undirected graph $G = (V, E)$, the maximum cut problem asks for a partition of the nodes V into two sets such that the number of edges between the two sets is maximized. This problem can be modeled as

$$\begin{aligned}
& \max \sum_{e \in E} y_e \\
& x_u + x_v + y_{\{u,v\}} \leq 2, & u, v \in V, \\
& -x_u - x_v + y_{\{u,v\}} \leq 0, & u, v \in V, \\
& x_v \in \{0, 1\}, & v \in V, \\
& y_e \in \{0, 1\}, & e \in E,
\end{aligned}$$

where x_v indicates to which of the two sets node $v \in V$ belongs to and y_e encodes whether edge $e \in E$ has its endpoints in different sets of the partition. This problem admits the reflection that simultaneously maps $x_v \mapsto 1 - x_v$ for all $v \in V$. Moreover, every automorphism π of G gives rise to a permutation symmetry that relabels the indices of x and y according to π .

Test Sets. We consider four test sets corresponding to the aforementioned problems. We refer to the test sets as “packing”, “kissing”, “energy”, and “maxcut”. The packing, kissing, and energy test set contain all instances with n objects in dimension d , where $(n, d) \in \{3, \dots, 14\} \times \{2, 3\}$. The maxcut test set contains an instance for each graph from the Color02 symposium².

²Obtained from <https://mat.tepper.cmu.edu/COLOR02/>. We removed the graph `games120` because the different settings hit the memory limit while solving the instance.

Table 1: Statistics on how many instances allow for a particular symmetry handling method.

test set	# instances		generators		row + column		row/column		simple
	total	sym.	sig.	unsig.	sig.	unsig.	sig.	unsig.	
structured instances:									
packing	24	24	24	24	23	0	0	0	0
kissing	24	24	24	24	23	0	0	0	0
energy	24	24	24	24	23	0	0	0	0
maxcut	119	119	119	94	0	0	0	0	25
benchmarking instances:									
miplib2017	1055	504	60	485	0	1	0	362	2
minplib	486	111	6	109	4	0	0	100	2
sat2002	819	292	282	125	0	0	9	78	54

Discussion Question (Q1). To answer (Q1), we have computed symmetries for all instances after presolving. Table 1 summarizes our results, where column “total” contains the total number of instances per test set and “sym.” the number of instances for which symmetries are detected. Column group “generators” lists the number of generators being signed (“sig.”) and unsigned (“unsig.”) permutations. Column groups “row + column” and “row/column” provide the number of instances containing row and column symmetries, and row or column symmetries, respectively. We distinguish whether the rows/columns can be reflected within the columns (“sig.” and “unsig.”). Finally, column “simple” reports on how many instances allow to handle symmetries via Inequality (5).

As Table 1 shows, we detect both permutation symmetries and signed permutation symmetries for all instances in the packing, kissing, and energy test set. Except for one instance per test set, row and column symmetries combined with column reflections are detected for all instances. The only exception are instances with specification $(n, d) = (3, 3)$, since both row and column symmetries have the same number of 2-cycles in this case. Our heuristic for detecting row and column symmetries thus cannot distinguish whether a permutation corresponds to a row or column permutation. Moreover, for each maxcut instance, we detect signed permutation symmetries. In particular, for 25 instances whose graphs are asymmetric, we also detect the applicability of Inequality (5). We conclude that our symmetry detection framework indeed reliably detects reflection symmetries.

Discussion Question (Q2). To answer (Q2), we have conducted experiments on the packing, kissing, and energy test set. We compare different ad-hoc approaches, which add symmetry handling inequalities directly to the problem formulation, with the more sophisticated techniques for row and column symmetries as described in Section 7.1. We refer to the latter setting as the “automatic” setting that is applied by SCIP. The remaining settings are inspired by the inequalities in [26] for the disk packing problem, also compare Proposition 6.1. Assuming that the corresponding variable matrix is $M \in \mathbb{R}^{p \times q}$, these settings are:

sym0: no symmetry handling inequalities are added;

sym1: add the inequalities $M_{1,1} \geq M_{1,2} \geq \dots \geq M_{1,q}$;

sym2: as sym1, additionally add $M_{1,1}, \dots, M_{1,q} \geq 0$;

sym3: as sym2, additionally add inequalities $M_{1,1} \geq M_{2,1} \geq \dots \geq M_{p,1}$;

sym4: using the notation of Proposition 6.1, add $M_{i,j} \geq 0$ for $j \in [q]$ and $i \in [n_j]$;

sym5: as sym4, also add $M_{n_j+1,1} \geq M_{n_j+2} \geq \dots \geq M_{n_{j-1},1}$ for $j \in [q] \cup \{0\}$;

sym6: combination of sym1 and sym5.

Table 2: Comparison of running times and primal-dual integrals for packing test set.

setting	dimension 2			dimension 3		
	# solved	time	primal-dual	# solved	time	primal-dual
sym0	4	605.02	14591.6	3	1928.92	43786.1
sym1	4	561.59	12891.3	3	1633.94	34028.4
sym2	6	457.49	8910.5	4	1308.92	21674.9
sym3	8	89.90	1238.4	5	639.00	10042.2
sym4	6	205.96	2150.2	5	815.80	8748.7
sym5	9	64.74	476.3	5	655.24	6845.1
sym6	9	58.60	584.9	5	508.59	5680.2
automatic	9	51.86	475.9	5	516.43	5266.5

Table 3: Comparison of running times and primal-dual integrals for kissing test set.

setting	dimension 2			dimension 3		
	# solved	time	primal-dual	# solved	time	primal-dual
sym0	4	381.16	7814.8	10	3.60	24.5
sym1	4	400.52	10813.1	10	3.87	26.8
sym2	5	342.43	9631.0	10	3.76	32.6
sym3	12	22.00	361.3	10	3.53	17.8
sym4	7	150.71	1885.1	10	3.69	36.0
sym5	11	32.96	546.5	10	4.79	44.5
sym6	11	24.88	330.7	10	3.80	21.9
automatic	12	10.57	170.0	10	3.77	21.6

Setting sym0 thus serves as a baseline to compare the different symmetry handling methods with an approach that does not handle symmetries. Moreover, note that settings sym1–sym3 and sym4–sym6 form two groups of settings, in which a higher number indicates to use more symmetry handling inequalities.

Tables 2–4 summarize our results, where columns “# solved”, “time”, and “primal-dual” provide the number of solved instances, the mean running time, and the mean primal-dual integral per test set, respectively. The primal-dual integral is a proxy for the speed of convergence of a branch-and-bound algorithm (roughly speaking, it measures the area between the primal and dual bounds plotted against time), see [6] for details. We included this measure, because we observed a very slow convergence of the dual bounds for some settings. Since optimal solutions are found usually rather quickly for these instances, a long running time but small integral indicates that almost optimal dual bounds are found early, but closing the gap completely is challenging. We rate an instance as solved when the primal-dual gap is below 0.1%.

The results show that symmetry handling is an important aspect for solving the three classes of problems more effectively. As expected, throughout all test sets (except for 3-dimensional kissing number problems, which are very easy), we observe that adding more symmetry handling inequalities to the problem formulation results in better running times. Methods sym3 and sym6 are thus the most effective settings within their groups. There is, however, no clear trend whether sym3 or sym6 is more effective. For example, for the kissing number problem and 2-dimensional energy problems, sym3 is faster, whereas sym6 dominates sym3 for the packing problem and 3-dimensional energy problems. In contrast, the automatic symmetry handling setting consistently dominates the other settings on 2-dimensional problems, and on the 3-dimensional packing problem it is almost the fastest approach. Compared to the best competitor, the automatic symmetry handling setting is 10.1% (2-dimensional packing), 52.0% (2-dimensional kissing), and 2.4% (2-dimensional energy) faster. On 3-dimensional energy problems, however, the automatic setting is slower than sym3 and sym6.

Investigating running times on a per-instance basis, see Appendix B for detailed numbers,

Table 4: Comparison of running times and primal-dual integrals for energy test set.

setting	dimension 2			dimension 3		
	# solved	time	primal-dual	# solved	time	primal-dual
sym0	3	1955.90	25054.0	0	7200.00	96744.4
sym1	3	1686.17	21190.0	1	6444.70	77109.6
sym2	3	1337.50	16811.4	1	5294.92	51906.2
sym3	6	346.94	2470.0	2	4398.29	33854.0
sym4	5	695.68	5615.2	1	5608.33	45764.8
sym5	7	402.21	2499.0	1	5607.37	40648.6
sym6	6	372.11	2384.9	2	3933.79	26150.2
automatic	6	338.69	1721.1	2	5275.07	32872.1

reveals that the automatic setting achieves the best results for the four aforementioned test sets for instances with a medium number of objects. For 2-dimensional packing problems with 9 and 10 balls it works substantially better than the competitors, whereas it works best for the 3-dimensional packing problem with 6 balls, the 2-dimensional kissing number problem with at least 9 spheres, and 9 points in the 2-dimensional energy problem. For instances with less objects, the automatic setting performs slightly worse than its competitors. Taking these observations into account might provide an explanation why the automatic setting performs worse than sym3 and sym6 on 3-dimensional energy instances: These instances are very challenging and the automatic setting as well as sym3 and sym6 only solve the instances with 3 and 4 points within the time limit of two hours. Extrapolating the results from the easier test sets thus could indicate that the automatic setting becomes more effective on instances with more points. To verify this hypothesis, we wanted to run the experiments with a higher time limit. However, all settings hit the memory limit before solving the more difficult instances. We thus could neither refute nor verify the hypothesis.

Besides the comparison of running times, note that, for the four easier test sets, the automatic setting has consistently a much smaller primal-dual integral than the best competitor. Thus, even if the running times are comparable for 3-dimensional packing and 2-dimensional energy problems, the corresponding primal-dual integral values are respectively 7.2% and 27.8% smaller for the automatic setting. The automatic setting hence tends to converge faster than the remaining methods.

In summary, the automatic setting is a competitive method that performs better than its competitors sym3 and sym6. While sym3 and sym6 can easily be incorporated into an optimization model by adding inequalities to a problem formulation, the automatic setting requires to implement the more sophisticated technique orbitopal reduction. The latter might be a technical burden preventing inexperienced practitioners to make use of this technique. Due to our symmetry detection framework for reflection symmetries combined with the heuristics for finding row and column symmetries, however, a solver can automatically detect that the sophisticated symmetry handling techniques are applicable. Users can thus benefit from powerful symmetry handling techniques that are built in a solver, and are not relying on handling reflection symmetries on their own.

Discussion of Results for Max-Cut. Table 5 summarizes our experiments for, i.a., the maxcut test set. For each test set, it mentions in parentheses the total number of instances of the test set as well as the number of instances that are solved by at least one setting. To compare different symmetry handling approaches, we use the mechanism described in Section 7.1 and possibly disable some techniques. The first four columns of Table 5 indicate which methods of this mechanism are used. A cross in column “sym.”, “row+col”, “refl.”, and “simpl.” respectively indicates that symmetry handling is active, row and column symmetries are detected, reflection symmetries are computed (if not, only permutation symmetries are computed), and Inequality (5) is applied when possible. To realize setting $(\times, \times, \times, \times)$, we disable lexicographic reduction to prevent that reflection symmetries are handled both by lexicographic reduction and (5). The remaining settings make use of lexicographic reduction if symmetry handling is enabled. Column groups “time” and “gap” report on the mean running times and average gap for all instances and the solvable

Table 5: Comparison of running times and gaps for symmetric instances.

setting				# solved	time		gap	
sym.	row+col	refl.	simp.		all	solved	all	solved
maxcut (118/41):								
				40	1220.19	42.61	35.37	0.05
×				40	1059.66	28.07	34.95	0.05
×	×			40	1058.43	27.97	34.94	0.05
×	×	×		41	1008.31	24.20	34.87	0.00
×	×	×	×	40	1070.91	28.96	35.33	0.08
miplib2017 (59/17):								
				15	2753.85	255.53	4416.48	588.24
×				14	2881.57	299.22	4588.75	590.82
×	×			14	2883.01	299.74	4588.75	590.82
×	×	×		17	2883.41	299.89	4414.99	0.00
×	×	×	×	16	2782.88	265.03	4414.41	1.95
minlplib (6/1):								
				1	1640.40	0.01	3380.72	0.00
×				1	1643.10	0.02	3379.81	0.00
×	×			1	1640.40	0.01	3375.20	0.00
×	×	×		1	1643.10	0.02	3370.42	0.00
×	×	×	×	1	1643.10	0.02	3370.42	0.00
sat2002 (282/172):								
				150	371.53	55.05	4680.85	1279.07
×				152	328.78	44.90	4609.93	1162.79
×	×			153	330.29	45.24	4574.47	1104.65
×	×	×		169	270.32	32.33	4007.09	174.42
×	×	×	×	153	354.20	50.84	4574.47	1104.65

instances. In contrast to the previous experiments, we rate an instance as solved if its gap is zero (up to numerical tolerances).

Comparing the running times of the different settings, we see that handling reflection symmetries using $(\times, \times, \times, \)$ is most effective. It solves one instances more than the remaining settings and reduces the running time in comparison to not detecting and handling reflection symmetries by 4.7% on all instances and 13.5% on the solvable instances. The setting $(\times, \times, \times, \times)$, which disables lexicographic reduction and instead applies (5) when it is applicable, still improves on the setting in which no symmetries are handled. But in comparison to the other symmetry handling approaches, it is the least performant. Since the motivation of this setting was to compare the effect of (5) with lexicographic reduction, we also separately considered the 25 instances for which (5) is applicable to see whether it has a positive effect there. None of these instances could be solved within the time limit though.

7.3 Results for Benchmarking Instances

Besides the structured instances discussed in the previous section, we also conducted experiments on general benchmarking instances. The test sets that we considered are all instances from MIPLIB2017 [17] and MINLPLIB [37], as well as the submitted instances of the SAT 2002 Competition [46]. To evaluate the impact of handling reflection symmetries, we removed all instances from these test sets for which no reflection symmetries could be detected. We refer to the corresponding test sets as miplib2017, minlplib, and sat2002, respectively.

In contrast to the structured instances, we cannot evaluate whether our framework reliably detects reflection symmetries for benchmarking instances. Our expectation was that reflection symmetries are rare for linear problems (miplib2017) and arise frequently for nonlinear problems

(minplib) and SAT problems. Indeed, as Table 1 shows, for 282 of the 819 instances from sat2002, we could detect reflection symmetries, whereas we could find only 60 instances from miplib2017 admitting reflection symmetries. Among the 486 instances from minplib, however, our framework could only detect 6 instances admitting reflection symmetries. This came as surprise to us, since MINLPLIB also contains instances corresponding to geometric packing problems (instances whose names start with “kall.”). Inspecting these instances revealed two explanations for not detecting the reflection symmetries. On the one hand, these instances already contain symmetry handling inequalities. On the other hand, in contrast to Example 1.1, the box in which the objects need to be placed is not fixed. Instead, one is looking for a box of minimal dimensions that can fit all objects. This is modeled asymmetrically by fixing the lower coordinate value and introducing a variable to model the upper coordinate value of each dimension. That is, although the real world problem admits reflection symmetries, the corresponding MINLP model is asymmetric.

In the following, we will therefore focus on the miplib2017 and sat2002 instances containing reflection symmetries, since the minplib test set is too small to draw reliable conclusions. The running times are summarized in Table 5. Note that the table reports only on 59 instances although Table 1 shows that there are 60 instances with reflection symmetries. To ensure a fair comparison of the different methods, however, we removed the instance “tokyometro” since all but one setting reached the memory limit.

Discussion of MIPLIB2017 For miplib2017, we observe that the $(\times, \times, \times, \times)$ setting performs best w.r.t. the number of solved instances. It can solve 17 instances, while just handling permutation symmetries can only solve 14 instances, and handling no symmetries at all solves 15 instances. Regarding the running time, however, $(\times, \times, \times, \times)$ and the settings only handling permutation symmetries perform equally and are on all instances 4.7% (on the solvable instances 17.1%) slower than not handling symmetries. It thus seems that handling reflection symmetries can help solving more instances, on average, however, it slows down the solving process. As such, it is not a surprise that the mean running time of $(\times, \times, \times, \times)$ is better than the one of $(\times, \times, \times, \times)$.

To understand why not handling symmetries performs better than handling symmetries, we compared the results for the 17 solvable instances for the setting in which no symmetries are handled and $(\times, \times, \times, \times)$. The following three observations could be made: (i) some instances are rather easy such that an improvement in running time is negligible; (ii) for the two instances that cannot be solved when not handling symmetries, also $(\times, \times, \times, \times)$ needed about 5900 s and 6400 s, respectively. That is, also when handling symmetries, the instances remain hard. (iii) The dual bound after presolving is (almost) optimal, i.e., it is sufficient to find an optimal solution. While the power of symmetry handling lies in pruning symmetric subproblems, which allows to more quickly improve the dual bound, it seems to hinder SCIP in finding feasible or optimal solutions. We conclude that, although handling symmetries on benchmarking instances has a positive effect in general [40], the characteristics of instances from MIPLIB2017 that admit reflection symmetries make symmetry handling less suited to enhance branch-and-bound for these instances.

The second question that arises is why the setting $(\times, \times, \times, \times)$ has the same mean running time as $(\times, \times, \times, \times)$ although it solves three more instances. Inspecting the symmetries that are found by the two different settings, we observed that the number of generators varies a lot between only detecting permutation symmetries and also reflection symmetries. For example, although the detected symmetry group for the instance `neos-3004026-krka` is larger when detecting reflection symmetries ($\approx 10^{91.5}$ group elements in comparison to $\approx 10^{90.9}$ for permutation symmetries), the number of generators we get from `bliss` is 35 for reflection symmetries and 64 for permutation symmetries. When handling symmetries via lexicographic reduction, we thus lose a lot of potential reductions when computing reflection symmetries. Moreover, for the instance `neos-780889`, we obtain the same number of generators corresponding to permutation symmetries; when handling reflection symmetries, however, we detect less column/row symmetries. That is, we miss the potential of specialized algorithms for column/row symmetries.

For the three additionally solved instances when handling reflection symmetries, we either find more generators (instance `icir07_tension`) or we detect more row/column symmetries (instances `lectsched-1` and `tanglegram4`). The explanation for the same mean running time thus indeed seems to be the variability in the generators returned by `bliss`.

Table 6: Comparison of running times and number of detected row/column symmetries for solvable sat2002 instances containing permutation symmetries.

setting				# solved	time	# row/column symmetries
sym.	row+col	refl.	simp.			
all instances (76):				72	63.10	1.00
×				74	39.67	14.87
×	×			74	39.80	14.87
×	×	×		74	50.57	7.01
×	×	×	×	74	53.28	7.01
feasible instances (27):				27	24.06	1.00
×				25	27.90	4.59
×	×			25	27.93	4.59
×	×	×		27	26.88	0.44
×	×	×	×	27	26.93	0.44
infeasible instances (49):				45	106.55	1.00
×				49	48.09	20.53
×	×			49	48.31	20.53
×	×	×		47	71.38	10.63
×	×	×	×	47	77.27	10.63

Discussion of SAT2002 On the sat2002 test set, the most effective setting is $(\times, \times, \times,)$. It solves 169 instances, and thus almost all solvable instances, within the time limit and improves upon only handling permutation symmetries by 17.8%. Taking Table 1 into account, this behavior is not surprising as at most 125 of the 292 reflection symmetric sat2002 instances contain permutation symmetries. That is, if reflection symmetries are not handled, a lot of instances become asymmetric.

To allow for a fair comparison between the different symmetry handling settings, we therefore also considered the subset of all solvable sat2002 instances that contain proper permutation symmetries. This results in 76 instances and the corresponding results are summarized in Table 6. On these instances, we observe that handling reflection symmetries on top of permutation symmetries decreases the performance by 27.5%, and this effect is even more pronounced on the infeasible instances, for which the running time increases by 48.4%. A possible explanation for this unexpected behavior is again the variance in the generators of the symmetry groups reported by `bliss`. While the mean number of row/column symmetries that are detected per instance are about 20.5 when only detecting permutation symmetries, the number of row/column symmetries drops to 10.6 when detecting reflection symmetries. That is, when detecting reflection symmetries, the potential of handling row/column symmetries by dedicated techniques cannot be exploited.

7.4 Conclusion and Outlook

In the introduction, we have formulated four main goals (G1)–(G4), which also could be achieved in this article: Our abstract framework of symmetry detection graphs turned out to be a flexible mechanism for detecting reflection symmetries in MINLP and beyond, cf. Goal (G1). Our open-source implementation could be used to detect reflection symmetries in many applications, cf. Goal (G2), and the numerical experiments showed that handling reflection symmetries can be crucial to accelerate branch-and-bound for specific applications, cf. Goal (G4). Although we devised methods for handling reflection symmetries, cf. Goal (G3), we noted that the performance improvement due to handling symmetries heavily depends on the structure of the detected symmetry groups. Handling reflection symmetries thus might slow down the solving process if this prevents our heuristics to detect row/column symmetries.

This latter observation opens directions for future research. As we noted in our experiments,

the generators of symmetry groups returned by symmetry detection tools such as `bliss` heavily depend on the structure of the symmetry detection graphs. Thus, based on the returned generators, our heuristics can fail to detect row and column symmetries. To circumvent this issue, it might be promising to develop alternative approaches for detecting row and column symmetries that depend less on the structure of generators. Ideally, one would use an exact mechanism to detect row/column symmetries, but detecting such symmetries is as hard as the graph isomorphism problem [7]. A possible future direction could thus be to exploit the specific structure of the symmetry detection graphs to solve the graph isomorphism problem.

Moreover, for MIPLIB2017, we noted that some problems benefit from not handling symmetries, because symmetry handling can hinder heuristics to find feasible solutions. A naive strategy for feasibility problems is thus to completely disable symmetry handling. For infeasible instances, however, this arguably slows down the solving process, since a larger search space needs to be explored until infeasibility is detected. Instead, it could be interesting to investigate means to benefit from handling symmetries in branch-and-bound, while removing the symmetry-based restrictions in heuristics.

Acknowledgements The author thanks Marc E. Pfetsch for very valuable discussions on the choice of the data structure for encoding symmetry detection graphs in SCIP as well as a thorough code review. This publication is part of the project “Local Symmetries for Global Success” with project number OCENW.M.21.299 which is financed by the Dutch Research Council (NWO).

References

- [1] Anders, M., Schweitzer, P.: Parallel computation of combinatorial symmetries. In: 29th Annual European Symposium on Algorithms, ESA 2021, September 6-8, 2021, Lisbon, Portugal (Virtual Conference), *LIPIcs*, vol. 204, pp. 6:1–6:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2021). DOI 10.4230/LIPIcs.ESA.2021.6. URL <https://doi.org/10.4230/LIPIcs.ESA.2021.6>
- [2] Anders, M., Schweitzer, P., Stieß, J.: Engineering a preprocessor for symmetry detection. CoRR **abs/2302.06351** (2023). DOI 10.48550/arXiv.2302.06351. URL <https://doi.org/10.48550/arXiv.2302.06351>
- [3] Babai, L., Luks, E.M.: Canonical labeling of graphs. In: Proceedings of the fifteenth annual ACM symposium on Theory of computing - STOC '83. ACM Press (1983). DOI 10.1145/800061.808746
- [4] Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J., Mahajan, A.: Mixed-integer nonlinear optimization. *Acta Numerica* **22**, 1–131 (2013). DOI 10.1017/S0962492913000032
- [5] Bendotti, P., Fouilhoux, P., Rottner, C.: Orbitopal fixing for the full (sub-)orbitope and application to the unit commitment problem. *Mathematical Programming* **186**, 337–372 (2021). DOI 10.1007/s10107-019-01457-1
- [6] Berthold, T.: Measuring the impact of primal heuristics. *Operations Research Letters* **41**(6), 611–614 (2013). DOI <https://doi.org/10.1016/j.orl.2013.08.007>. URL <https://www.sciencedirect.com/science/article/pii/S0167637713001181>
- [7] Berthold, T., Pfetsch, M.E.: Detecting orbitopal symmetries. In: B. Fleischmann, K.H. Borgwardt, R. Klein, A. Tuma (eds.) *Operations Research Proceedings 2008*, pp. 433–438. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
- [8] Bessiere, C., Hebrard, E., Hnich, B., Walsh, T.: The complexity of global constraints. In: *Proceedings of the 19th National Conference on AI* (2004)
- [9] Bödi, R., Herr, K., Joswig, M.: Algorithms for highly symmetric linear and integer programs. *Mathematical Programming* **137**(1), 65–90 (2013). DOI 10.1007/s10107-011-0487-6. URL <http://dx.doi.org/10.1007/s10107-011-0487-6>

- [10] Bolusani, S., Besançon, M., Bestuzheva, K., Chmiela, A., Dionísio, J., Donkiewicz, T., van Doornmalen, J., Eiffler, L., Ghannam, M., Gleixner, A., Graczyk, C., Halbig, K., Hedtke, I., Hoen, A., Hojny, C., van der Hulst, R., Kamp, D., Koch, T., Kofler, K., Lentz, J., Manns, J., Mexi, G., Mühmer, E., Pfetsch, M.E., Schlösser, F., Serrano, F., Shinano, Y., Turner, M., Vigerske, S., Weninger, D., Xu, L.: The SCIP Optimization Suite 9.0 (2024)
- [11] Cohen, J.S.: Computer algebra and symbolic computation: mathematical methods. AK Peters, Natick, Massachusetts (2003)
- [12] Costa, A., Hansen, P., Liberti, L.: On the impact of symmetry-breaking constraints on spatial branch-and-bound for circle packing in a square. *Discrete Applied Mathematics* **161**(1), 96–106 (2013). DOI <https://doi.org/10.1016/j.dam.2012.07.020>. URL <https://www.sciencedirect.com/science/article/pii/S0166218X12002855>
- [13] van Doornmalen, J., Hojny, C.: A unified framework for symmetry handling (2023). DOI 10.48550/arXiv.2211.01295
- [14] van Doornmalen, J., Hojny, C.: Efficient propagation techniques for handling cyclic symmetries in binary programs. *INFORMS Journal on Computing* **0**(0), null (2024). DOI 10.1287/ijoc.2022.0060
- [15] Flener, P., Frisch, A.M., Hnich, B., Kiziltan, Z., Miguel, I., Pearson, J., Walsh, T.: Breaking row and column symmetries in matrix models. In: P. Van Hentenryck (ed.) *Principles and Practice of Constraint Programming - CP 2002*, pp. 462–477. Springer Berlin Heidelberg, Berlin, Heidelberg (2002)
- [16] Friedman, E.J.: Fundamental domains for integer programs with symmetries. In: A. Dress, Y. Xu, B. Zhu (eds.) *Combinatorial Optimization and Applications, Lecture Notes in Computer Science*, vol. 4616, pp. 146–153. Springer Berlin Heidelberg (2007). DOI 10.1007/978-3-540-73556-4_17
- [17] Gleixner, A., Hendel, G., Gamrath, G., Achterberg, T., Bastubbe, M., Berthold, T., Christophel, P.M., Jarck, K., Koch, T., Linderoth, J., Lübbecke, M., Mittelmann, H.D., Ozyurt, D., Ralphs, T.K., Salvagnin, D., Shinano, Y.: MIPLIB 2017: Data-Driven Compilation of the 6th Mixed-Integer Programming Library. *Mathematical Programming Computation* pp. 443–490 (2021). DOI 10.1007/s12532-020-00194-3. URL <https://doi.org/10.1007/s12532-020-00194-3>
- [18] Hojny, C.: Supplementary material for the article “Detecting and handling reflection symmetries in mixed-integer (nonlinear) programming”. <https://doi.org/10.5281/zenodo.11189482>
- [19] Hojny, C.: Packing, partitioning, and covering symresacks. *Discrete Applied Mathematics* **283**, 689–717 (2020). DOI 10.1016/j.dam.2020.03.002
- [20] Hojny, C.: Polynomial size IP formulations of knapsack may require exponentially large coefficients. *Operations Research Letters* **48**(5), 612–618 (2020). DOI <https://doi.org/10.1016/j.orl.2020.07.013>. URL <http://www.sciencedirect.com/science/article/pii/S0167637720301103>
- [21] Hojny, C., Pfetsch, M.E.: Polytopes associated with symmetry handling. *Mathematical Programming* **175**, 197–240 (2019). DOI 10.1007/s10107-018-1239-7
- [22] Junttila, T., Kaski, P.: Conflict propagation and component recursion for canonical labeling. In: A. Marchetti-Spaccamela, M. Segal (eds.) *Theory and Practice of Algorithms in (Computer) Systems – First International ICST Conference, TAPAS 2011, Rome, Italy, April 18–20, 2011. Proceedings, Lecture Notes in Computer Science*, vol. 6595, pp. 151–162. Springer (2011). DOI 10.1007/978-3-642-19754-3_16

- [23] Junntila, T., Kaski, P.: Conflict propagation and component recursion for canonical labeling. In: A. Marchetti-Spaccamela, M. Segal (eds.) *Theory and Practice of Algorithms in (Computer) Systems – First International ICST Conference, TAPAS 2011, Rome, Italy, April 18–20, 2011. Proceedings, Lecture Notes in Computer Science*, vol. 6595, pp. 151–162. Springer (2011). DOI [10.1007/978-3-642-19754-3_16](https://doi.org/10.1007/978-3-642-19754-3_16)
- [24] Kaibel, V., Peinhardt, M., Pfetsch, M.E.: Orbitopal fixing. *Discrete Optimization* **8**(4), 595–610 (2011). DOI <http://dx.doi.org/10.1016/j.disopt.2011.07.001>
- [25] Kaibel, V., Pfetsch, M.E.: Packing and partitioning orbitopes. *Mathematical Programming* **114**(1), 1–36 (2008). DOI [10.1007/s10107-006-0081-5](https://doi.org/10.1007/s10107-006-0081-5)
- [26] Khajavirad, A.: Packing circles in a square: a theoretical comparison of various convexification techniques (2017). URL <https://optimization-online.org/?p=14462>
- [27] Korte, B., Vygen, J.: *Combinatorial Optimization: Theory and Algorithms*, 6 edn. Springer, Heidelberg (2018)
- [28] Liberti, L.: Automatic generation of symmetry-breaking constraints. In: *Combinatorial optimization and applications, Lecture Notes in Computer Science*, vol. 5165, pp. 328–338. Springer, Berlin (2008). DOI [10.1007/978-3-540-85097-7_31](https://doi.org/10.1007/978-3-540-85097-7_31)
- [29] Liberti, L.: Symmetry in mathematical programming. In: J. Lee, S. Leyffer (eds.) *Mixed Integer Nonlinear Programming, IMA Series*, vol. 154, pp. 263–283. Springer New York (2011). DOI [10.1007/978-1-4614-1927-3_9](https://doi.org/10.1007/978-1-4614-1927-3_9)
- [30] Liberti, L.: Reformulations in mathematical programming: automatic symmetry detection and exploitation. *Mathematical Programming* **131**(1-2), 273–304 (2012). DOI [10.1007/s10107-010-0351-0](https://doi.org/10.1007/s10107-010-0351-0). URL <http://dx.doi.org/10.1007/s10107-010-0351-0>
- [31] Liberti, L., Ostrowski, J.: Stabilizer-based symmetry breaking constraints for mathematical programs. *Journal of Global Optimization* **60**, 183–194 (2014)
- [32] Linderoth, J., Núñez Ares, J., Ostrowski, J., Rossi, F., Smriglio, S.: Orbital conflict: Cutting planes for symmetric integer programs. *INFORMS Journal on Optimization* **3**(2), 139–153 (2021). DOI [10.1287/ijoo.2019.0044](https://doi.org/10.1287/ijoo.2019.0044)
- [33] Margot, F.: Pruning by isomorphism in branch-and-cut. *Mathematical Programming* **94**(1), 71–90 (2002). DOI [10.1007/s10107-002-0358-2](https://doi.org/10.1007/s10107-002-0358-2)
- [34] Margot, F.: Exploiting orbits in symmetric ILP. *Mathematical Programming* **98**(1-3), 3–21 (2003). DOI [10.1007/s10107-003-0394-6](https://doi.org/10.1007/s10107-003-0394-6)
- [35] Margot, F.: Symmetry in integer linear programming. In: M. Jünger, T.M. Liebling, D. Naddef, G.L. Nemhauser, W.R. Pulleyblank, G. Reinelt, G. Rinaldi, L.A. Wolsey (eds.) *50 Years of Integer Programming*, pp. 647–686. Springer (2010)
- [36] McKay, B.D., Piperno, A.: Practical graph isomorphism, II. *Journal of Symbolic Computation* **60**, 94–112 (2014). DOI <https://doi.org/10.1016/j.jsc.2013.09.003>
- [37] Minlplib: A library of mixed-integer and continuous nonlinear programming instances. <http://minlplib.org/index.html>
- [38] Ostrowski, J.: Symmetry in integer programming. PhD dissertation, Lehigh University (2009)
- [39] Ostrowski, J., Linderoth, J., Rossi, F., Smriglio, S.: Orbital branching. *Mathematical Programming* **126**(1), 147–178 (2011). DOI [10.1007/s10107-009-0273-x](https://doi.org/10.1007/s10107-009-0273-x)
- [40] Pfetsch, M.E., Rehn, T.: A computational comparison of symmetry handling methods for mixed integer programs. *Mathematical Programming Computation* **11**(1), 37–93 (2019). DOI [10.1007/s12532-018-0140-y](https://doi.org/10.1007/s12532-018-0140-y)

- [41] Puget, J.F.: Automatic Detection of Variable and Value Symmetries, pp. 475–489. Springer Berlin Heidelberg, Berlin, Heidelberg (2005). DOI 10.1007/11564751_36. URL http://dx.doi.org/10.1007/11564751_36
- [42] Saff, E., Kuijlaars, A.: Distributing many points on a sphere. *The Mathematical Intelligencer* **19**, 5–11 (1997)
- [43] Sakallah, K.A.: Handbook of Satisfiability, Editors: Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh, chap. Symmetry and Satisfiability. IOS Press (2021)
- [44] Salvagnin, D.: A dominance procedure for integer programming. Master’s thesis, University of Padova, Padova, Italy (2005)
- [45] Salvagnin, D.: Symmetry breaking inequalities from the Schreier-Sims table. In: W.J. van Hove (ed.) *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pp. 521–529. Springer International Publishing (2018). DOI 10.1007/978-3-319-93031-2_37
- [46] Sat 2002 competition: problem instances. <https://www.cs.ubc.ca/~hoos/SATLIB/Benchmarks/SAT/New/Competition-02/sat-2002-beta.tgz>
- [47] Szabó, P.G., Markót, M.C., Csendes, T.: *Global Optimization in Geometry — Circle Packing into the Square*, pp. 233–265. Springer US, Boston, MA (2005). DOI 10.1007/0-387-25570-2_9
- [48] Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming* **106**, 25–57 (2006). DOI <https://doi.org/10.1007/s10107-004-0559-y>
- [49] Zhu, W.: Unsolvability of some optimization problems. *Applied Mathematics and Computation* **174**(2), 921–926 (2006). DOI <https://doi.org/10.1016/j.amc.2005.05.025>

A Overview of Important Functions to Apply Our Symmetry Detection Framework

This appendix provides an overview of the most important functions needed to extend an SDG within a SCIP symmetry detection callback. Since our implementation of SDGs allows for four different types of nodes, we have different functions for adding these nodes:

SCIPaddSymgraphOpnode()	adds an operator node to an SDG;
SCIPaddSymgraphValnode()	adds a numerical value node to an SDG;
SCIPaddSymgraphConsnode()	adds a constraint node to an SDG.

Recall that we do not allow to add variable nodes to an SDG, because SCIP ensures that every SDG contains all necessary variable nodes. Instead, the indices of variable nodes can be accessed via the functions

SCIPgetSymgraphVarnodeidx()	returns the index of the node corresponding to a given variable;
SCIPgetSymgraphNegatedVarnodeidx()	returns the index of the node corresponding to a negated/reflected variable.

To add edges to a graph, the function

SCIPaddSymgraphEdge()	adds an edge between two existing nodes of an SDG
-----------------------	---

can be used.

To simplify the usage of SDGs, we also provide two functions that add gadgets for certain variable structures to an SDG:

<code>SCIPextendPermsymDetectionGraphLinear()</code>	adds a gadget for a linear expression $a^\top x + b$ to an SDG;
<code>SCIPaddSymgraphVarAggregation()</code>	adds a gadget for aggregated variables to an SDG.

The second function has been introduced, since we require that no aggregated or fixed variables are present in an SDG.

B Detailed Numerical Results

In this appendix, we provide detailed numerical results for the tested problem classes. Tables 7–12 report on the running times and primal-dual integrals for each instance of the 2- and 3-dimensional packing, kissing number, and energy problems that we discussed in Section 7.2. The number of items corresponds to the number of balls, spheres, and points in these respective problems, whereas the settings refer to the settings sym0–sym6 and the automatic setting as described in Section 7.2.

Table 7: Running times and primal-dual integrals for packing test set and dimension 2.

# items	setting							
	sym0	sym1	sym2	sym3	sym4	sym5	sym6	auto.
running time in seconds:								
3	0.12	0.12	0.11	0.05	0.09	0.07	0.09	0.09
4	2.84	1.94	0.67	0.30	0.44	0.38	0.29	0.28
5	0.79	0.59	0.38	0.17	0.30	0.31	0.22	0.21
6	43.09	25.56	7.41	0.61	0.50	0.40	0.70	0.68
7	7200.00	7200.00	6606.63	22.81	201.83	16.16	11.01	16.23
8	7200.00	7200.00	4352.59	10.14	70.32	22.82	6.68	12.14
9	7200.00	7200.00	7200.00	644.78	7200.00	249.99	365.62	103.95
10	7200.00	7200.00	7200.00	267.77	7200.00	173.67	153.78	61.42
11	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
12	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
13	7200.00	7200.00	7200.00	7200.00	7200.00	358.37	351.27	301.84
14	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
primal-dual integral:								
3	5	5	6	2	3	2	3	4
4	35	28	16	11	9	9	8	8
5	29	23	19	10	14	14	12	11
6	802	527	151	22	12	9	26	24
7	120 899	109 585	31 757	210	941	109	133	123
8	145 839	118 007	52 138	285	1489	227	90	120
9	234 515	269 194	152 641	6449	67 037	1741	2826	667
10	266 727	224 203	229 408	5454	73 878	2567	1027	754
11	355 432	344 742	334 154	154 772	102 571	66 800	74 434	50 360
12	361 376	353 135	351 864	199 890	128 913	20 578	77 572	68 018
13	368 337	354 376	353 221	190 939	98 344	7071	8769	6855
14	408 953	408 172	405 794	277 031	200 833	46 888	128 374	105 141

Table 8: Running times and primal-dual integrals for packing test set and dimension 3.

# items	setting							
	sym0	sym1	sym2	sym3	sym4	sym5	sym6	auto.
running time in seconds:								
3	2.09	0.93	0.35	0.29	0.22	0.25	0.27	0.41
4	1.49	0.85	0.29	0.29	0.31	0.31	0.29	0.28
5	7200.00	7200.00	5921.24	380.15	626.00	414.19	67.69	73.69
6	6663.82	1961.23	341.28	10.73	34.86	13.46	8.40	5.87
7	7200.00	7200.00	7200.00	631.20	2443.29	645.28	287.70	395.11
8	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
9	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
10	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
11	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
12	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
13	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
14	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
primal-dual integral:								
3	17	11	9	6	5	6	7	9
4	19	12	5	11	10	10	9	8
5	52 376	22 897	5366	437	566	421	86	88
6	48 670	15 778	2794	210	474	128	126	94
7	154 538	164 221	110 088	6448	10 830	3605	2901	2755
8	259 241	231 341	176 139	133 512	126 275	120 202	106 350	106 084
9	281 270	271 892	224 430	154 967	92 668	66 342	79 984	69 743
10	319 452	304 708	285 611	206 888	154 403	158 242	138 025	127 942
11	342 112	334 613	314 555	251 442	144 333	128 962	117 638	115 709
12	361 785	345 513	334 208	267 405	171 725	163 406	166 023	145 259
13	362 833	355 450	341 441	277 927	143 289	173 997	127 984	115 555
14	373 245	388 226	343 286	299 399	227 500	207 009	208 462	153 976

Table 9: Running times and primal-dual integrals for kissing test set and dimension 2.

# items	setting							
	sym0	sym1	sym2	sym3	sym4	sym5	sym6	auto.
running time in seconds:								
3	0.02	0.03	0.08	0.01	0.06	0.02	0.04	0.04
4	0.07	0.18	0.16	0.14	0.10	0.15	0.16	0.17
5	0.06	0.35	0.28	0.17	0.39	0.29	0.17	0.17
6	0.16	0.48	0.47	0.23	0.31	0.61	0.61	0.26
7	7200.00	7200.00	1136.23	2.36	22.42	3.62	1.75	2.72
8	7200.00	7200.00	7200.00	5.89	340.06	18.33	6.34	6.11
9	7200.00	7200.00	7200.00	11.85	451.55	9.87	7.22	4.50
10	7200.00	7200.00	7200.00	51.76	7200.00	86.73	54.52	17.89
11	7200.00	7200.00	7200.00	175.53	7200.00	355.42	126.20	29.58
12	7200.00	7200.00	7200.00	448.92	7200.00	1408.36	1578.97	380.60
13	7200.00	7200.00	7200.00	1928.51	7200.00	3137.21	2976.38	116.21
14	7200.00	7200.00	7200.00	5501.44	7200.00	7200.00	7200.00	849.15
primal-dual integral:								
3	2	3	7	1	6	2	2	3
4	5	14	16	8	9	11	9	10
5	6	30	24	9	31	29	14	14
6	9	42	45	20	25	47	33	17
7	38 404	21 079	5254	99	209	159	54	74
8	245 267	213 439	101 851	264	1873	651	149	201
9	313 339	306 493	286 624	342	1978	193	159	205
10	392 407	414 595	404 572	1598	94 434	1503	1095	406
11	491 427	491 463	484 365	4463	104 650	5516	1994	682
12	527 099	527 103	524 806	9572	239 255	13 014	18 173	5704
13	555 106	555 107	555 094	45 741	152 492	34 012	29 414	1369
14	577 444	578 765	577 441	130 808	177 820	328 974	121 905	10 812

Table 10: Running times and primal-dual integrals for kissing test set and dimension 3.

# items	setting							
	sym0	sym1	sym2	sym3	sym4	sym5	sym6	auto.
running time in seconds:								
3	0.02	0.02	0.03	0.02	0.05	0.02	0.01	0.02
4	0.02	0.03	0.04	0.03	0.05	0.03	0.02	0.02
5	0.04	0.04	0.04	0.01	0.07	0.02	0.02	0.02
6	0.04	0.03	0.03	0.03	0.07	0.02	0.03	0.02
7	0.05	0.05	0.06	0.04	0.08	0.04	0.03	0.04
8	0.04	0.04	0.07	0.03	0.06	0.03	0.05	0.08
9	0.07	0.08	0.07	0.04	0.08	1.37	0.07	0.03
10	0.09	0.09	0.08	0.05	0.13	0.07	0.73	0.05
11	0.10	0.13	0.54	0.07	0.09	1.97	0.12	0.92
12	0.10	1.09	0.14	0.06	0.15	2.08	0.20	0.07
13	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
14	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
primal-dual integral:								
3	2	2	3	2	5	2	1	2
4	2	3	4	3	5	3	2	2
5	4	4	4	1	7	2	2	2
6	4	3	3	3	7	2	3	2
7	5	4	6	4	8	4	3	4
8	4	4	7	3	6	3	5	8
9	7	8	7	4	8	137	7	3
10	9	9	8	5	13	7	18	5
11	10	13	54	7	9	197	9	33
12	10	22	14	6	15	193	11	7
13	61 672	61 411	61 412	61 568	61 612	61 583	62 150	62 493
14	92 105	92 101	92 102	93 511	92 106	92 116	93 969	92 138

Table 11: Running times and primal-dual integrals for energy test set and dimension 2.

# items	setting							
	sym0	sym1	sym2	sym3	sym4	sym5	sym6	auto.
running time in seconds:								
3	2.17	0.96	0.37	0.14	0.31	0.30	0.20	0.14
4	28.92	14.75	4.69	0.90	3.64	1.87	1.36	1.35
5	637.64	330.00	80.48	3.96	6.50	3.74	2.06	2.00
6	7200.00	7200.00	7200.00	20.58	169.95	36.29	19.36	19.48
7	7200.00	7200.00	7200.00	131.63	1669.89	171.30	95.21	80.70
8	7200.00	7200.00	7200.00	733.38	7200.00	2358.56	7200.00	7200.00
9	7200.00	7200.00	7200.00	7200.00	7200.00	3555.87	3074.32	1258.01
10	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
11	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
12	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
13	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
14	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
primal-dual integral:								
3	6	4	3	3	2	3	5	3
4	59	30	22	5	14	12	11	11
5	1048	733	175	38	25	21	17	13
6	27 577	16 940	15 275	105	339	95	67	60
7	103 265	91 918	71 028	558	3115	460	300	221
8	165 382	159 741	138 685	3004	61 183	5533	8552	4083
9	209 730	204 553	183 465	13 291	78 012	9411	7622	3704
10	248 606	244 166	225 948	56 374	138 046	64 303	60 746	38 110
11	268 921	265 619	246 456	101 186	161 925	98 961	98 287	76 904
12	284 270	286 673	276 693	153 360	201 235	146 922	145 253	126 222
13	303 207	301 486	290 631	178 830	210 680	163 230	162 411	123 375
14	313 677	311 820	308 313	214 843	243 370	206 209	200 767	165 437

Table 12: Running times and primal-dual integrals for energy test set and dimension 3.

# items	setting							
	sym0	sym1	sym2	sym3	sym4	sym5	sym6	auto.
running time in seconds:								
3	7200.00	1904.11	179.29	73.14	358.43	357.69	24.35	772.85
4	7200.00	7200.00	7200.00	1889.64	7200.00	7200.00	1448.27	1602.78
5	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
6	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
7	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
8	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
9	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
10	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
11	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
12	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
13	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
14	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00	7200.00
primal-dual integral:								
3	1528	447	44	23	80	80	11	186
4	14 737	6781	2051	491	2658	1855	385	408
5	77 151	59 093	36 746	12 087	16 408	12 026	5531	5184
6	121 289	104 487	84 450	41 473	61 835	46 654	31 565	31 476
7	157 158	143 920	126 967	86 664	100 511	88 191	72 522	71 813
8	178 762	168 522	156 057	120 629	133 847	120 473	109 227	108 159
9	195 888	186 304	172 755	146 157	134 636	128 682	116 954	115 462
10	208 562	199 569	187 568	164 420	158 820	151 279	141 834	140 935
11	219 253	213 879	201 490	185 224	174 839	168 161	159 821	159 343
12	225 733	223 219	210 340	192 819	188 728	182 468	175 968	174 244
13	233 884	229 828	221 233	205 331	197 518	190 162	185 491	184 832
14	240 182	238 349	227 048	216 176	209 512	202 497	198 844	198 483